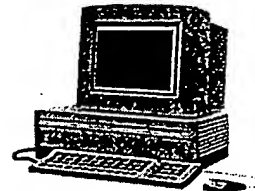


BioTech-Chem Library

Search Results

Feedback Form (Optional)



Scientific & Technical Information C

The search results generated for your recent request are attached. If you have any questions or comments (compliments or complaints) about the scope or the results of the search, please contact *the BioTech-Chem searcher* who conducted the search *or contact*:

Mary Hale, Supervisor, 308-4:
CM-1 Room 1E01

Voluntary Results Feedback Form

➤ *I am an examiner in Workgroup:* (Example: 1610)

➤ *Relevant prior art found, search results used as follows:*

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ *Relevant prior art not found:*

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Search results were not useful in determining patentability or understanding the invention.

Other Comments:

Drop off completed forms at the Circulation Desk CM-1, or send to Mary Hale, CM1-1E01 or mary.hale@uspto.gov

THIS PAGE BLANK (USPTO)

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

017N: 2d start

84524

Access DB#

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Celine Qian Examiner #: 75770 Date: 1/16/03
Art Unit: 1636 Phone Number 306-0283 Serial Number: 09/822935
Mail Box and Bldg/Room Location: 11E012 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Regulatory sequence of mouse villin gene - used in transgenesis
Inventors (please provide full names): Daniel Pinto et al.

Earliest Priority Filing Date: 2/19/1998

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please send for SEQ ID NO: 1 & a oligo search of SEQ ID NO.

CRFEE

Call
2/19/03

RECEIVED
JAN 16 2003
STIC/Biotech

Edward Han
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

THIS PAGE BLANK (USPTO)

PN W020034492-A1.
XX 15-JUN-2000.
PD
XX 09-DEC-1998; 98WO-EP08009.
PF
XX 09-DEC-1998; 98WO-EP08009.
PR
XX (CNRS) CENT NAT RECH SCI.
PA (CURIE-) INST CURIE.
XX
XX Plato D, Robine S, Jaisser F, Louvard D;
PI
XX WPI: 2000-423433/36.
DR
XX Novel nucleotide sequence derived from mouse villin gene for targeted
PT expression of transgenes in immature and differentiated epithelial
PT cells of intestine or urogenital tracts
PS
XX Claim 3; Fig 6; 54pp; English.
XX
CC The present sequence comprises the murine villin gene regulatory
CC region, the first intron and part of the first two exons. It has been
CC shown that this region directs the expression of the villin gene in the
CC intestine and uro-genital tracts, and thus could be used in a fusion
CC gene to direct expression of exogenous genes in these areas. This could
CC be used, for example, to create a mouse model for colorectal cancer.
XX
SQ Sequence 8995 BP: 2275 A; 2105 C; 2258 G; 2357 T; 0 other:

Query Match 100.0%; Score 8995; DB 21; Length 8995;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGGGTGCACCAAGCACTGTGGTCCACGACTGGGGAGGTGGAGGAGGCTCA 60
DB 1 GATCGGGTGCACCAAGCACTGTGGTCCACGACTGGGGAGGTGGAGGAGGCTCA 60
QY 61 GAAGTTTAAGGTCATCTTGGTTACATAGCAAGTTTCAGCAGCTTCAGCTACATGAAA 120
DB 61 GAAGTTTAAGGTCATCTTGGTTACATAGCAAGTTTCAGCAGCTTCAGCTACATGAAA 120
QY 121 CCTTGGTTGTTGTTGTTGTTGTTTAAAGCATTAATAAATTAATACATAAGAGGTTGG 180
DB 121 CCTTGGTTGTTGTTGTTGTTGTTTAAAGCATTAATAAATTAATACATAAGAGGTTGG 180
QY 181 CAGTGTGGCAGACACCTTAATTCAGATATTCAGAGGCGAAGACAGCAGATCTCTGT 240
DB 181 CAGTGTGGCAGACACCTTAATTCAGATATTCAGAGGCGAAGACAGCAGATCTCTGT 240
QY 241 GAGTTCGAAGTCAGGCTGTGTGCAAAAGCTAGTTCAGAGTGGCAAGGGCTACACAGAGA 300
DB 241 GAGTTCGAAGTCAGGCTGTGTGCAAAAGCTAGTTCAGAGTGGCAAGGGCTACACAGAGA 300
QY 301 AACCTTGTCTCATATAAACCAAGTAGTAGTAGTAGTAATGCCATAGAGAAAATTGGA 360
DB 301 AACCTTGTCTCATATAAACCAAGTAGTAGTAGTAGTAATGCCATAGAGAAAATTGGA 360
QY 361 GTCCATTAGAGATGAGCACCCTATAGATGATTTCTTGACCCAGGTAAAGTAAATGTCA 420
DB 361 GTCCATTAGAGATGAGCACCCTATAGATGATTTCTTGACCCAGGTAAAGTAAATGTCA 420
QY 421 TGGGGAAGGGATGGAGCTGTCTAGATTAAAAAGCTGTGAGGGATGCCATTCTGAA 480
DB 421 TGGGGAAGGGATGGAGCTGTCTAGATTAAAAAGCTGTGAGGGATGCCATTCTGAA 480
QY 481 TTTGATTCATATGAAGAGCTGATAGGCCCCAAGAGAAGTGAAGTGGACTGTGACT 540
DB 481 TTTGATTCATATGAAGAGCTGATAGGCCCCAAGAGAAGTGAAGTGGACTGTGACT 540
QY 541 GAAGACGTGACGGCTTTAAACACTGCACTTATAAACAACACTTATAAACAACACTGGCAGG 600
DB 541 GAAGACGTGACGGCTTTATAAACAACACTTATAAACAACACTTATAAACAACACTGGCAGG 600

QY 601 CGTTCAGGTTTGAAGATCATCTTTCAAAACACAGAACAGAAAGTGCTGCTGCTCAGC 660
DB 601 CGTTCAGGTTTGAAGATCATCTTTCAAAACACAGAACAGAAAGTGCTGCTGCTCAGC 660
QY 661 GTAGGAGCACTGCGTCGAGAGAGTGTATTTAGTGAAGCTACCTTCACAAATATCTTT 720
DB 661 GTAGGAGCACTGCGTCGAGAGAGTGTATTTAGTGAAGCTACCTTCACAAATATCTTT 720
QY 721 GCACATTACATATACAGTGTGCAAAATGTGCTAAGTCCCTAGTCCACAGATGGCTTTACA 780
DB 721 GCACATTACATATACAGTGTGCAAAATGTGCTAAGTCCCTAGTCCACAGATGGCTTTACA 780
QY 781 CTGCTTTCTGCTTCCCATCTGTTGACATTTGTGCAAGAACGAAATAGAAATGTGGCT 840
DB 781 CTGCTTTCTGCTTCCCATCTGTTGACATTTGTGCAAGAACGAAATAGAAATGTGGCT 840
QY 841 ATTTATTTGTGTGTGAGAGACACATCCAGGGCTTTTCATTTTCAGGCACATGCTTTAC 900
DB 841 ATTTATTTGTGTGTGAGAGACACATCCAGGGCTTTTCATTTTCAGGCACATGCTTTAC 900
QY 901 TAACGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTATTTTGTGT 960
DB 901 TAACGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTATTTTGTGT 960
QY 961 GCATGAGTAGGATGTATAGATGTATAGAGATGATAGAGATGATAGAGATGATAGAGAT 1020
DB 961 GCATGAGTAGGATGTATAGATGTATAGAGATGATAGAGATGATAGAGATGATAGAGAT 1020
QY 1021 ATCATTTGAGATCCCGACAGAGTGAAGTACCGAGCGTTGTAGTTGTTATGTGGAGCTG 1080
DB 1021 ATCATTTGAGATCCCGACAGAGTGAAGTACCGAGCGTTGTAGTTGTTATGTGGAGCTG 1080
QY 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGACAGCAGTGGCTTTAACATGGACACAGCTCT 1140
DB 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGACAGCAGTGGCTTTAACATGGACACAGCTCT 1140
QY 1141 CTAGGCTTAAGATTAATCTTTGTTTAAATAATATATTTCTACGGGGGTGTGGTGGC 1200
DB 1141 CTAGGCTTAAGATTAATCTTTGTTTAAATAATATATTTCTACGGGGGTGTGGTGGC 1200
QY 1201 ACACGCCCTTAATCCACAGCCTTGAGAGGCTGAGAGTGTAGGAATTTATACACAGCCAG 1260
DB 1201 ACACGCCCTTAATCCACAGCCTTGAGAGGCTGAGAGTGTAGGAATTTATACACAGCCAG 1260
QY 1261 CTGGGCTGCAAGCTTGGCCCTGTTTTTTTTTTTTTTTTTTTCTTATGTGCACTGGTCTTA 1320
DB 1261 CTGGGCTGCAAGCTTGGCCCTGTTTTTTTTTTTTTTTTTTTCTTATGTGCACTGGTCTTA 1320
QY 1321 CCTGGCTATGTCCGTGCAAGGCTGTGAGATCCCTTGAGACTGGAGTTAAAGACAGTTG 1380
DB 1321 CCTGGCTATGTCCGTGCAAGGCTGTGAGATCCCTTGAGACTGGAGTTAAAGACAGTTG 1380
QY 1381 TGATCACGCTCCGTTACAGATGCTGGAATTTGAACCCAGGCTGCTTACAGAACAGCC 1440
DB 1381 TGATCACGCTCCGTTACAGATGCTGGAATTTGAACCCAGGCTGCTTACAGAACAGCC 1440
QY 1441 AGTGTCTTAATCTCTGAGCCACCCCTCAACCCCTGCTTTAGAGACATCTTAACCTTTTG 1500
DB 1441 AGTGTCTTAATCTCTGAGCCACCCCTCAACCCCTGCTTTAGAGACATCTTAACCTTTTG 1500
QY 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTTACCAAGTGTGTGCTGCTAGACATCA 1560
DB 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTTACCAAGTGTGTGCTGCTAGACATCA 1560
QY 1561 CTGAGCCCGTACCCACACAGACTAGTGTAGATGTTTAAAGGCAACACTTAACAATGACA 1620
DB 1561 CTGAGCCCGTACCCACACAGACTAGTGTAGATGTTTAAAGGCAACACTTAACAATGACA 1620
QY 1621 ATAGTGTGATAGAGTTTGAATATAGTCTTGAGCTATTGGTTAGGCTGACCTTTGTGTC 1680
DB 1621 ATAGTGTGATAGAGTTTGAATATAGTCTTGAGCTATTGGTTAGGCTGACCTTTGTGTC 1680

| | | | |
|----|------|--|------|
| OY | 1661 | TTACACATGCTGCTGGAGAAATATGAAAAATGAAAGACTTGAGCTGTAGTCTGCTTGGAAACCACA | 1740 |
| OY | 1661 | TTTACATGCTGCTGGAGAAATATGAAAAATGAAAGACTTGAGCTGTAGTCTGCTTGGAAACCACA | 1740 |
| Db | 1661 | TTTACATGCTGCTGGAGAAATATGAAAAATGAAAGACTTGAGCTGTAGTCTGCTTGGAAACCACA | 1740 |
| OY | 1741 | GAGCAGCGCGAAGACCCACCTCTGAAAGTGTCTCTGAGCTTCACATACAACTTCACAT | 1800 |
| Db | 1741 | GAGCAGCGCGAAGACCCACCTCTGAAAGTGTCTCTGAGCTTCACATACAACTTCACAT | 1800 |
| OY | 1801 | AATAGTTACATGATTAATTAATTAATAGTAAATYCTTTTAAAAAGTATATGTTGGAGGGA | 1860 |
| Db | 1801 | AATAGTTACATGATTAATTAATTAATAGTAAATYCTTTTAAAAAGTATATGTTGGAGGGA | 1860 |
| OY | 1861 | GAGATGCGCTCAGCTCCAGAGACACTTGCTGCTCTTGACAGAGGACCTTGATTCAGTTCC | 1920 |
| Db | 1861 | GAGATGCGCTCAGCTCCAGAGACACTTGCTGCTCTTGACAGAGGACCTTGATTCAGTTCC | 1920 |
| OY | 1921 | AGGACTATATATGAGGCTCTCAGCCACTCTGTAAATCCATTCACAGAGGTTCCACACCTT | 1980 |
| Db | 1921 | AGGACTATATATGAGGCTCTCAGCCACTCTGTAAATCCATTCACAGAGGTTCCACACCTT | 1980 |
| OY | 1981 | CTTCTGGCCTCCACAGGACACACATACATAGTACACAGACATACATGACAGGCAAAACACC | 2040 |
| Db | 1981 | CTTCTGGCCTCCACAGGACACACATACATAGTACACAGACATACATGACAGGCAAAACACC | 2040 |
| OY | 2041 | CATACACACATTAATTAATTAAGSAAACTTPAAAGGTGCATGTGTTGGTTPAACATTTGCT | 2100 |
| Db | 2041 | CATACACACATTAATTAATTAAGSAAACTTPAAAGGTGCATGTGTTGGTTPAACATTTGCT | 2100 |
| OY | 2101 | TACACATGCTGATTTGAAGACATGTATCACAGGACACACCTCAAGAGGATCTGGGCTGGAG | 2160 |
| Db | 2101 | TACACATGCTGATTTGAAGACATGTATCACAGGACACACCTCAAGAGGATCTGGGCTGGAG | 2160 |
| OY | 2161 | AGATGCGCTCAGCGGTTAAAGACACTGACTGCTCTTCCGAAAGAAAGTCTCTGAGTTCAAT | 2220 |
| Db | 2161 | AGATGCGCTCAGCGGTTAAAGACACTGACTGCTCTTCCGAAAGAAAGTCTCTGAGTTCAAT | 2220 |
| OY | 2221 | CCTAGCAACACATGCTGCTCACAACCATTCATAATGAGATCTGACACCCCTCTTGCT | 2280 |
| Db | 2221 | CCTAGCAACACATGCTGCTCACAACCATTCATAATGAGATCTGACACCCCTCTTGCT | 2280 |
| OY | 2281 | GCACTCTAAGACAGCTGACAGCTACAGTGTACTTAAGATATCTAATTAATTAATCTTTT | 2340 |
| Db | 2281 | GCACTCTAAGACAGCTGACAGCTACAGTGTACTTAAGATATCTAATTAATTAATCTTTT | 2340 |
| OY | 2341 | TTTAAAAAAATGAAGAGGATCTGAGACACACTCAAAAGAGATTAGACAGTCACTCAG | 2400 |
| Db | 2341 | TTTAAAAAAATGAAGAGGATCTGAGACACACTCAAAAGAGATTAGACAGTCACTCAG | 2400 |
| OY | 2401 | GGTGATTATCTATCTGAGAGTCTTTTCCCTTCCGCTTGCAACTGGGTGGACAGC | 2460 |
| Db | 2401 | GGTGATTATCTATCTGAGAGTCTTTTCCCTTCCGCTTGCAACTGGGTGGACAGC | 2460 |
| OY | 2461 | CCCTTTTCAATTCACAAGAACGGGTCTCTACATTATTTCTGAACAAACAGCACTGCAGT | 2520 |
| Db | 2461 | CCCTTTTCAATTCACAAGAACGGGTCTCTACATTATTTCTGAACAAACAGCACTGCAGT | 2520 |
| OY | 2521 | ATGTTTACTCTCTTGGCTGACTATGAGACAGCGGACGCGCGCCGACACACACACAC | 2580 |
| Db | 2521 | ATGTTTACTCTCTTGGCTGACTATGAGACAGCGGACGCGCGCCGACACACACACAC | 2580 |
| OY | 2581 | ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAGGTCA | 2640 |
| Db | 2581 | ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAGGTCA | 2640 |
| OY | 2641 | AGAAAGAGGCTGCTTCAACACATCTTCATCTTCCCTCAAAAGAGACACAGATTC | 2700 |
| Db | 2641 | AGAAAGAGGCTGCTTCAACACATCTTCATCTTCCCTCAAAAGAGACACAGATTC | 2700 |
| OY | 2701 | AAGGTGGCAGAAATCTTACAGGGGGCAGAGGCAAGGAGGGGGAACAGGCAATGGTTTCC | 2760 |
| Db | 2701 | AAGGTGGCAGAAATCTTACAGGGGGCAGAGGCAAGGAGGGGGAACAGGCAATGGTTTCC | 2760 |
| OY | 2761 | AGAGACCTTACAGCAGAGGGGACGAAAGGACAGATCCCAAGGCTCCAGGGCAGGAGAGGTGGAG | 2820 |

| | | | |
|----|------|--|------|
| Db | 2761 | AGAAACCTACACGCGAGAGGGGACGAGACAGCAATCCCCAGGTCCAGGGGACAGGGAGTGGAGG | 2820 |
| Oy | 2821 | CCCTGTTTCCGAGGAGAAAGCCAGGCGGACAAACAGGGTTCAAGGCACACAGGTTTATGGCA | 2880 |
| Db | 2821 | CCCTGTTTCCGAGGAGAAAGCCAGGCGGACAAACAGGGTTCAAGGCACACAGGTTTATGGCA | 2880 |
| Oy | 2881 | GCTCATAAAGTGGAGGTCGTGGCTCACTCAGAAAGGAGGAAGAGGAAAGGCGCCCTTGT | 2940 |
| Db | 2881 | GCTCATAAAGTGGAGGTCGTGGCTCACTCAGAAAGGAGGAAGAGGAAAGGCGCCCTTGT | 2940 |
| Oy | 2941 | GCCCACTGAGCGAGGGTCATGCTGAGTAGGAGAGATCTCGAGGGTGGCCAGGAGCCCCAC | 3000 |
| Db | 2941 | GCCCACTGAGCGAGGGTCATGCTGAGTAGGAGAGATCTCGAGGGTGGCCAGGAGCCCCAC | 3000 |
| Oy | 3001 | CTGCTGTGCCCAAGGGAAGCCCAAGTGTGAACCTGAGCTTGAGGTGTCGATGTCACGTA | 3060 |
| Db | 3001 | CTGCTGTGCCCAAGGGAAGCCCAAGTGTGAACCTGAGCTTGAGGTGTCGATGTCACGTA | 3060 |
| Oy | 3061 | CAAGACCCCAAGAGTCCCTACTCCATCCCATCCAGTGGCCCCCGCCCGCCACACCCCA | 3120 |
| Db | 3061 | CAAGACCCCAAGAGTCCCTACTCCATCCCATCCAGTGGCCCCCGCCCGCCACACCCCA | 3120 |
| Oy | 3121 | CCCCGACTCCCGTGCACATTCTCTAGGGCTGAGGGTGGCAGGCCCTGATGGGGTTGC | 3180 |
| Db | 3121 | CCCCGACTCCCGTGCACATTCTCTAGGGCTGAGGGTGGCAGGCCCTGATGGGGTTGC | 3180 |
| Oy | 3181 | CTACCTCAGGTGAGGCCGACAGGTCCTAGCCGGAAGTGCACCCCATCTCCGAAAGTCGACGA | 3240 |
| Db | 3181 | CTACCTCAGGTGAGGCCGACAGGTCCTAGCCGGAAGTGCACCCCATCTCCGAAAGTCGACGA | 3240 |
| Oy | 3241 | GCCAAAGGGCGGGGACACAGGCAAGCTCAGAGCTGTGACAGCTGTGAGGCTCAGATTGCC | 3300 |
| Db | 3241 | GCCAAAGGGCGGGGACACAGGCAAGCTCAGAGCTGTGACAGCTGTGAGGCTCAGATTGCC | 3300 |
| Oy | 3301 | AGGACCTGGGACCTACCTTCCACCACCCCCCATCTCTCTGAGGGCCCTATCTTCC | 3360 |
| Db | 3301 | AGGACCTGGGACCTACCTTCCACCACCCCCCATCTCTCTGAGGGCCCTATCTTCC | 3360 |
| Oy | 3361 | CTTATATGCTGAAGAGATTCCTGAGGGGGGGGGGGTGGTGGTGAAGACAAAGCTGTGCG | 3420 |
| Db | 3361 | CTTATATGCTGAAGAGATTCCTGAGGGGGGGGGGGTGGTGGTGAAGACAAAGCTGTGCG | 3420 |
| Oy | 3421 | GTCCTCTGACAGCGACTTGGCACACACTTCTCTAAGATCTCCAGGTGGTGGCTCTTCC | 3480 |
| Db | 3421 | GTCCTCTGACAGCGACTTGGCACACACTTCTCTAAGATCTCCAGGTGGTGGCTCTTCC | 3480 |
| Oy | 3481 | CAGACAGGTAAAGCAATTTGGGTGGGGACACATGTTGACACAGGTGGTGGAGGGGACAG | 3540 |
| Db | 3481 | CAGACAGGTAAAGCAATTTGGGTGGGGACACATGTTGACACAGGTGGTGGAGGGGACAG | 3540 |
| Oy | 3541 | GGTCTTGTCTTCTCTCTGAGAGGCTGTGCTTCTCTGAGACCTTGGTATAAAGTTGGGGG | 3600 |
| Db | 3541 | GGTCTTGTCTTCTCTCTGAGAGGCTGTGCTTCTCTGAGACCTTGGTATAAAGTTGGGGG | 3600 |
| Oy | 3601 | TGAGGTAAAGTGCTCTGAACCTCTGAAGAAGCAACAACAGCAGCGCTTGGGCT | 3660 |
| Db | 3601 | TGAGGTAAAGTGCTCTGAACCTCTGAAGAAGCAACAACAGCAGCGCTTGGGCT | 3660 |
| Oy | 3661 | TCAATGAAGGAAGTTCACAGACCCCTTTCCTGTAGTACCTGACCTTGCTTCACTGTGTAG | 3720 |
| Db | 3661 | TCAATGAAGGAAGTTCACAGACCCCTTTCCTGTAGTACCTGACCTTGCTTCACTGTGTAG | 3720 |
| Oy | 3721 | ATTCCCTGAGGACCAAGTGGCTCTCTGGGACTACAGATTTCTACAAATTAATCAGGACGT | 3780 |
| Db | 3721 | ATTCCCTGAGGACCAAGTGGCTCTCTGGGACTACAGATTTCTACAAATTAATCAGGACGT | 3780 |
| Oy | 3781 | CCTGAGACTTGGACCTCGGCTGTATTTAATACACTCTCTGAGGCTCAATTTCTGTGT | 3840 |
| Db | 3781 | CCTGAGACTTGGACCTCGGCTGTATTTAATACACTCTCTGAGGCTCAATTTCTGTGT | 3840 |
| Oy | 3841 | TCATGCTTACACATCTGAATGGTTTCTTTGTGTACCAATTCCTCTACACTCTGTGGA | 3900 |

Db 3841 TCAGTCTTACACATCTGAATGTTCTTGTGTACCATTCCTCCGACACTCCCTGGGA 3900
Oy 3901 GGTGCTATCCTGGACATGTATCCTGGGATGTAGCTGCAGCCACCGAGAGAGGGGG 3960
Db 3901 GGTGCTATCCTGGACATGTATCCTGGGATGTAGCTGCAGCCACCGAGAGAGGGGG 3960
Oy 3961 AGAGTCAGAGAGCTGTCTCAGGAGGCTATAGGAGCTGCAGCATCACCCCTTCTCTGAAT 4020
Db 3961 AGAGTCAGAGAGCTGTCTCAGGAGGCTATAGGAGCTGCAGCATCACCCCTTCTCTGAAT 4020
Oy 4021 GGGCCCTCCATTTTTCGGTTACCATGTATTTATATCAGAGGGGAGGAAAGCCA 4080
Db 4021 GGGCCCTCCATTTTTCGGTTACCATGTATTTATATCAGAGGGGAGGAAAGCCA 4080
Oy 4081 AACCTGCCAGAAAGTTGGGACTCAGACCAAGGTTATGCTCGAAATCCCCCTG 4140
Db 4081 AACCTGCCAGAAAGTTGGGACTCAGACCAAGGTTATGCTCGAAATCCCCCTG 4140
Oy 4141 TCACCTTGAAGTTGGAGAAATGCTCTGGGGCTTCAGAGCTTGTGTTAGCAGAGAGGT 4200
Db 4141 TCACCTTGAAGTTGGAGAAATGCTCTGGGGCTTCAGAGCTTGTGTTAGCAGAGAGGT 4200
Oy 4201 ATCTTTGTATAGGCAATGACCTAGTCTATGCTGTACTACATTCCTGCTCAATTAAG 4260
Db 4201 ATCTTTGTATAGGCAATGACCTAGTCTATGCTGTACTACATTCCTGCTCAATTAAG 4260
Oy 4261 CTGGAAGTAAACCCAGGAGGAGGAGGATTCCTACAGTTGATCCCAAGAACCA 4320
Db 4261 CTGGAAGTAAACCCAGGAGGAGGAGGATTCCTACAGTTGATCCCAAGAACCA 4320
Oy 4321 AGACAGTATATGCAAGATAGTAGTGGGAGAGAAAGTAAACCCCAAG 4380
Db 4321 AGACAGTATATGCAAGATAGTAGTGGGAGAGAAAGTAAACCCCAAG 4380
Oy 4381 GGGCAGAGGTTCCCTTCCCTAGTTCACAATGCAATGCAATGCAATGCAATGCAATG 4440
Db 4381 GGGCAGAGGTTCCCTTCCCTAGTTCACAATGCAATGCAATGCAATGCAATGCAATG 4440
Oy 4441 TGAGTGTAGTATGCAAGATAGTATGCTGTATGCTGTATGCTGTATGCTGTATG 4500
Db 4441 TGAGTGTAGTATGCAAGATAGTATGCTGTATGCTGTATGCTGTATGCTGTATG 4500
Oy 4501 TTGGAGGCTGAAAGCAGAGATGCTATATGTTTGAAGCCAGCTGACTATAGAGCA 4560
Db 4501 TTGGAGGCTGAAAGCAGAGATGCTATATGTTTGAAGCCAGCTGACTATAGAGCA 4560
Oy 4561 GACTTGTCTTTAAGAAAAATGAAGCCCAAGGCTGTGGCAACGCCCTTAATCCA 4620
Db 4561 GACTTGTCTTTAAGAAAAATGAAGCCCAAGGCTGTGGCAACGCCCTTAATCCA 4620
Oy 4621 GCACCTTGGAGGAGCAGAGCAGATTTCTGATTTCAAGGCGCAGCTGTATAGAGT 4680
Db 4621 GCACCTTGGAGGAGCAGAGCAGATTTCTGATTTCAAGGCGCAGCTGTATAGAGT 4680
Oy 4681 GAGTTCCAGGAGAGCCAGGCTACACAGAAACCCTGTTGAAAAACCAAGAAACAA 4740
Db 4681 GAGTTCCAGGAGAGCCAGGCTACACAGAAACCCTGTTGAAAAACCAAGAAACAA 4740
Oy 4741 AACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCA 4800
Db 4741 AACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCA 4800
Oy 4801 TAGGCTGTCTGTCTAGTGTAGTGTGGGACTCAGACTATATATATATATAGGCC 4860
Db 4801 TAGGCTGTCTGTCTAGTGTAGTGTGGGACTCAGACTATATATATATATAGGCC 4860
Oy 4861 TTTTATACATGTGTACAGAGAGAAAGTTTCACTGTGGACACAGTGGAGCCCTGAGA 4920
Db 4861 TTTTATACATGTGTACAGAGAGAAAGTTTCACTGTGGACACAGTGGAGCCCTGAGA 4920
Oy 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGGAAGGCTTCTTGAAGAAAGTGTCTCCGAT 4980
Db 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGGAAGGCTTCTTGAAGAAAGTGTCTCCGAT 4980

Oy 4981 CAGACTACTGTTCTAGAGGAGAGAGAGGTTGGAAGATGTTGGTGACAGACACTT 5040
Db 4981 CAGACTACTGTTCTAGAGGAGAGAGAGGTTGGAAGATGTTGGTGACAGACACTT 5040
Oy 5041 GGAACAGAGAGCAGAGAGGAGGAGCATCCAGATTTCTGAACATGTAGCTGACTTTTGGT 5100
Db 5041 GGAACAGAGAGCAGAGAGGAGGAGCATCCAAAGATTTCTGAACATGTAGCTGACTTTTGGT 5100
Oy 5101 TCTCTGGGTGACAGTGTCCCCAGGAGTATGGGCTGTAGAAAGGGGACCAAGGCTGAGCC 5160
Db 5101 TCTCTGGGTGACAGTGTCCCCAGGAGTATGGGCTGTAGAAAGGGGACCAAGGCTGAGCC 5160
Oy 5161 AATGAGTTCAGGTTAGGAGACATCCAGCCAGGCTGCTGCTGGCAAGCTAAAGATG 5220
Db 5161 AATGAGTTCAGGTTAGGAGACATCCAGCCAGGCTGCTGCTGGCAAGCTAAAGATG 5220
Oy 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGAGAGAGAGAGTGAAGATCCCTCTA 5280
Db 5221 AGAGCCCTCTAACCCCTCCCTGAAGTTTAGGGAGAGAGAGAGTGAAGATCCCTCTA 5280
Oy 5281 GGTGAGAGAGAGTATGCTGCTGACCAACATGCTTAGAGAGAGACAGTGGAGCAG 5340
Db 5281 GGTGAGAGAGAGTATGCTGCTGACCAACATGCTTAGAGAGAGACAGTGGAGCAG 5340
Oy 5341 TTACCCCTCAGAACAGGCAATCCCTTGGCTTAAGAGAGGCTGGGCCCTTCTGTTT 5400
Db 5341 TTACCCCTCAGAACAGGCAATCCCTTGGCTTAAGAGAGGCTGGGCCCTTCTGTTT 5400
Oy 5401 AAGAACTTACTTTTCTTACAGAGAGAGCAGCAAGCCTTGTCCCTCCCTGTTGGTCAA 5460
Db 5401 AAGAACTTACTTTTCTTACAGAGAGAGCAGCAAGCCTTGTCCCTCCCTGTTGGTCAA 5460
Oy 5461 TAAACCCCTGCTGTAAACATAGTTATTTTACTGTACAGTTGCTCAGAGACGTCCA 5520
Db 5461 TAAACCCCTGCTGTAAACATAGTTATTTTACTGTACAGTTGCTCAGAGACGTCCA 5520
Oy 5521 TCTGTAGACCTCTGCTCTTAACCTCAGCAAGTATGGCCCATTCCTCACCAGAAAG 5580
Db 5521 TCTGTAGACCTCTGCTCTTAACCTCAGCAAGTATGGCCCATTCCTCACCAGAAAG 5580
Oy 5581 TGCAGAGAGAGCCTTAGAGAAAGGTTAAAGATTAACAAAGATGGCCAAATTAACAA 5640
Db 5581 TGCAGAGAGAGCCTTAGAGAAAGGTTAAAGATTAACAAAGATGGCCAAATTAACAA 5640
Oy 5641 ACTACTATCCTTTGTAACCAATTTGCTTGTGAACAGAGAGGAGTGTGAGTGTAT 5700
Db 5641 ACTACTATCCTTTGTAACCAATTTGCTTGTGAACAGAGAGGAGTGTGAGTGTAT 5700
Oy 5701 GGT 5760
Db 5701 GGT 5760
Oy 5761 CTTGGGAGACTTTTATCATCTTAAGAAATATCTGATTTTGGCCCAATGCAAGAGGATAT 5820
Db 5761 CTTGGGAGACTTTTATCATCTTAAGAAATATCTGATTTTGGCCCAATGCAAGAGGATAT 5820
Oy 5821 TGGGAGAGTCAAGGCTGTGCAACACAGTAAAGTGGCCCAAGTGTGAGTGGCTGAA 5880
Db 5821 TGGGAGAGTCAAGGCTGTGCAACACAGTAAAGTGGCCCAAGTGTGAGTGGCTGAA 5880
Oy 5881 TCACCAAGGAGGAGCTATCAGATGAGAGACATCAAGATTAACCAACCTGTGG 5940
Db 5881 TCACCAAGGAGGAGCTATCAGATGAGAGACATCAAGATTAACCAACCTGTGG 5940
Oy 5941 GCTCAGAAAGGAGTTCACAGAGATTAAGGCCAAGCCTATTTATTCACAGACATGAC 6000
Db 5941 GCTCAGAAAGGAGTTCACAGAGATTAAGGCCAAGCCTATTTATTCACAGACATGAC 6000
Oy 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGAGAGATGGGGCTGTAGTGTGGACACC 6060
Db 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGAGAGATGGGGCTGTAGTGTGGACACC 6060

| | | | |
|----|------|--|------|
| OY | 6081 | TGACCTTGCACTTATTAGTACTAAGGCCAAAGAGCAGTGTACAGAGGGTGACTGGGTCCTA | 6120 |
| Db | 6081 | TGACCTTGCACTTATTAGTACTAAGGCCAAAGAGCAGTGTACAGAGGGTGACTGGGTCCTA | 6120 |
| OY | 6121 | CTCAGCTTGGAGCGAGCAGCTGGAGAATGGGATGACCCTCATCTCATGTGAGAGGGGCTTAG | 6180 |
| Db | 6121 | CTCAGCTTGGAGCGAGCAGCTGGAGAATGGGATGACCCTCATCTCATGTGAGAGGGGCTTAG | 6180 |
| OY | 6181 | CACGACAGGTTACAAGTGTCCCTGTGTCTCATGTGCAGGATTCCTGGCCAGTTTTCCAAG | 6240 |
| Db | 6181 | CACGACAGGTTACAAGTGTCCCTGTGTGTCTCATGTGCAGGATTCCTGGCCAGTTTTCCAAG | 6240 |
| OY | 6241 | GACTTAAGGACTCATCTCTGGTGGAAAACAAAGTATCCAGCCCTTAAGCCCCATTTTGGTCT | 6300 |
| Db | 6241 | GACTTAAGGACTCATCTCTGGTGGAAAACAAAGTATCCAGCCCTTAAGCCCCATTTTGGTCT | 6300 |
| OY | 6301 | AATTAAATCAGAACCCCCTGGGGATGCGAGGCTCGACACACAGACTTTTAAAAAGCTC | 6360 |
| Db | 6301 | AATTAAATCAGAACCCCCTGGGGATGCGAGGCTCGACACACAGACTTTTAAAAAGCTC | 6360 |
| OY | 6361 | CCAGGTGATTTCTGATCGACAGCTGGAAACAAACAGACTTCAGGTTCAAACAGAAAAGAGC | 6420 |
| Db | 6361 | CCAGGTGATTTCTGATCGACAGCTGGAAACAAACAGACTTCAGGTTCAAACAGAAAAGAGC | 6420 |
| OY | 6421 | AAACCTTAGGGAAAGCTTGGGATGGGGAGCCCTTTCGAGGCCAGTAGATGAGAGCTGTT | 6480 |
| Db | 6421 | AAACCTTAGGGAAAGCTTGGGATGGGGAGCCCTTTCGAGGCCAGTAGATGAGAGCTGTT | 6480 |
| OY | 6481 | AGCAGTGGTGGACAGCTTCTCTGCGCTGTCATATACCTATCCATCCATCCATCCAT | 6540 |
| Db | 6481 | AGCAGTGGTGGACAGCTTCTCTGCGCTGTCATATACCTATCCATCCATCCATCCATCCAT | 6540 |
| OY | 6541 | ACACCCACCCATCCATTTATGCAACCCATCCCTTCATCCATCCATCCATCCATCCATCCAC | 6600 |
| Db | 6541 | ACACCCACCCATCCATTTATGCAACCCATCCCTTCATCCATCCATCCATCCATCCATCCAC | 6600 |
| OY | 6601 | CCACGACTCCATCCAAACCTTCCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 6660 |
| Db | 6601 | CCACGACTCCATCCAAACCTTCCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 6660 |
| OY | 6661 | CATTATATCCAAAGAGAACGATATGPTACTAAAGTGGGAATTAATTAATTTTATGA | 6720 |
| Db | 6661 | CATTATATCCAAAGAGAACGATATGPTACTAAAGTGGGAATTAATTAATTTTATGA | 6720 |
| OY | 6721 | AGCTCTGTGATTTGACTGATTTGCAATGATGTGACAGAGTGCATATCCACAGCACAGCTG | 6780 |
| Db | 6721 | AGCTCTGTGATTTGACTGATTTGCAATGATGTGACAGAGTGCATATCCACAGCACAGCTG | 6780 |
| OY | 6781 | TGGCAATCGGAGAAAGGTTTTGGGTGTTGTTTTCTCTCCACCGCTGTGGGTTCTGGGGA | 6840 |
| Db | 6781 | TGGCAATCGGAGAAAGGTTTTGGGTGTTGTTTTCTCTCCACCGTGTGGGTTCTGGGGA | 6840 |
| OY | 6841 | TTGGAATCCAAATTAATCGGCGTGGGTGGCAGGTGTCCTTACACCGAGCCATTTTGTGACA | 6900 |
| Db | 6841 | TTGGAATCCAAATTAATCGGCGTGGGTGGCAGGTGTCCTTACACCGAGCCATTTTGTGACA | 6900 |
| OY | 6901 | CATCATTTATTATTTAGAAGACCTTTATGTATAGTCCAGAGCTGAGCCTCAACCTTGCTATGCG | 6960 |
| Db | 6901 | CATCATTTATTATTTAGAAGACCTTTATGTATAGTCCAGAGCTGAGCCTCAACCTTGCTATGCG | 6960 |
| OY | 6961 | CCACGAGTGAACCTTAACTCTGCTCTTCAGAGCTCCACCCAGTAGTCAAGGTTTACAGGT | 7020 |
| Db | 6961 | CCACGAGTGAACCTTAACTCTGCTCTTCAGAGCTCCACCCAGTAGTCAAGGTTTACAGGT | 7020 |
| OY | 7021 | GTTCAACTGGTGAATGCCCTTAAATCCAGACCTCTGTGGGGGGGGGGGAGGCGGATC | 7080 |
| Db | 7021 | GTTCAACTGGTGAATGCCCTTAAATCCAGACCTCTGTGGGGGGGGGGGAGGCGGATC | 7080 |
| OY | 7081 | CCTGAGTTGGAGGCACTTTGGTCTCAGAGTTTACAGGTTACCTGGGCTATACAGGGAA | 7140 |
| Db | 7081 | CCTGAGTTGGAGGCACTTTGGTCTCAGAGTTTACAGGTTTACAGGTTTACAGGGAA | 7140 |
| OY | 7141 | ACCCTATCCCAACAAACAAACAAACAAACAAATAATTTGTCTCAATTAATTCACAGGA | 7200 |

| | | | |
|----|------|---|------|
| Dd | 7141 | ACCCATATCCCAAAACAACAAACAAACAAAAAATATTCTGTSCAATAATATCACAGAGA | 7200 |
| Oy | 7201 | TTAGAGGATATATAGTAGGGAGTAgGGGCTGGTGAGGAAGCATGCTTTCTTTGTATT | 7260 |
| Dd | 7201 | TTAGAGGATATATAGTAGGGAGTAgGGGCTGGTGAGGAAGCATGCTTTCTTTGTATT | 7260 |
| Oy | 7261 | ATAATAGTAAGTACTCACAAGATGCATTATCTATCTATCTATCTATCTATCTATCTATC | 7320 |
| Dd | 7261 | ATAATAGTAAGTACTCACAAGATGCATTATCTATCTATCTATCTATCTATCTATCTATC | 7320 |
| Oy | 7321 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC | 7380 |
| Dd | 7321 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC | 7380 |
| Oy | 7381 | CTGCTTTGACGCTGAATGTCCTATTTCTGGGGCACTCTTACCCTAGTTGGGTTT | 7440 |
| Dd | 7381 | CTGCTTTGACGCTGAATGTCCTATTTCTGGGGCACTCTTACCCTAGTTGGGTTT | 7440 |
| Oy | 7441 | ACCAACACCCAGACATTTATTTTTATTTTGTATTATTTATATCTAGAGCTCAGGGTG | 7500 |
| Dd | 7441 | ACCAACACCCAGACATTTATTTTTATTTTGTATTATTTATATCTAGAGCTCAGGGTG | 7500 |
| Oy | 7501 | GGACTCAGGGCTTGTGTGCTATGCTTAAGCAAGCTCTCGCCACAGAGCTGAGCTCCAGTCC | 7560 |
| Dd | 7501 | GGACTCAGGGCTTGTGTGCTATGCTTAAGCAAGCTCTCGCCACAGAGCTGAGCTCCAGTCC | 7560 |
| Oy | 7561 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCTATTTGCCAGGCTATGTAGCTCTCTC | 7620 |
| Dd | 7561 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCTATTTGCCAGGCTATGTAGCTCTCTC | 7620 |
| Oy | 7621 | CACCTCCAGTTTCCAGCAGCTTCTGTGCTATCCGAGTGGGGGGAACACTCTGTGCTCACCA | 7680 |
| Dd | 7621 | CACCTCCAGTTTCCAGCAGCTTCTGTGCTATCCGAGTGGGGGGAACACTCTGTGCTCACCA | 7680 |
| Oy | 7681 | GTCGCCCTGTTCCTGCTTTTATGAGGGGGCCCCCTTATCTTATGACAGATTGTGATG | 7740 |
| Dd | 7681 | GTCGCCCTGTTCCTGCTTTTATGAGGGGGCCCCCTTATCTTATGACAGATTGTGATG | 7740 |
| Oy | 7741 | GGATGCGTTCCTGTGATTTCTTTTATGAGGGGGCCCCCTTATCTTATGACAGATTGTG | 7800 |
| Dd | 7741 | GGATGCGTTCCTGTGATTTCTTTTATGAGGGGGCCCCCTTATCTTATGACAGATTGTG | 7800 |
| Oy | 7801 | GGCCATGTGTCACGCTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT | 7860 |
| Dd | 7801 | GGCCATGTGTCACGCTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT | 7860 |
| Oy | 7861 | TGTGTGATTAACCACTTCTGTTTCAATTACTGATGAAATTTGGCCCCACCCCAC | 7920 |
| Dd | 7861 | TGTGTGATTAACCACTTCTGTTTCAATTACTGATGAAATTTGGCCCCACCCCAC | 7920 |
| Oy | 7921 | CCTTTTTTTTTATTTGACACAAGGCTTCTGTGCTATCTTGSAACTTGGCTGCT | 7980 |
| Dd | 7921 | CCTTTTTTTTTATTTGACACAAGGCTTCTGTGCTATCTTGSAACTTGGCTGCT | 7980 |
| Oy | 7981 | GAGCTCACTCTGTGACACAGGCTGTAGGCTGCTCTCCACTTTTGACACTCTGTGAAC | 8040 |
| Dd | 7981 | GAGCTCACTCTGTGACACAGGCTGTAGGCTGCTCTCCACTTTTGACACTCTGTGAAC | 8040 |
| Oy | 8041 | AGAGTAGCCATGAACCTTCAAAGACAATTTTGTGTTTGGTTTGTTTTACATTTGTGTG | 8100 |
| Dd | 8041 | AGAGTAGCCATGAACCTTCAAAGACAATTTTGTGTTTGGTTTGTTTTACATTTGTGTG | 8100 |
| Oy | 8101 | TGTATGGCTGTATATGTGCACTGTTTGTGTCTTCAGGTGCTCACAIGTGTATACCTGTGTG | 8160 |
| Dd | 8101 | TGTATGGCTGTATATGTGCACTGTTTGTGTCTTCAGGTGCTCACAIGTGTATACCTGTGTG | 8160 |
| Oy | 8161 | TGGACACAGACAACCAACCGATGTGGCATTCCTCAGATCTAGTGAATATATGA | 8220 |
| Dd | 8161 | TGGACACAGACAACCAACCGATGTGGCATTCCTCAGATCTAGTGAATATATGA | 8220 |
| Oy | 8221 | TGTATTTATGATGTTTATTTTATGTTGCTCCCAAGTATGACGATTTTGTGTGAGTTTTCAC | 8280 |

```

Db 8221 TGTATTTATGATGTTTATTTAGTGTGCCCAAGTATGACAGGATTTTGTGAGATTTTCAC 8280
Oy 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTAGAGCAATGCCCTCA 8340
Db 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTAGAGCAATGCCCTCA 8340
Oy 8341 CTTCGATGAGCAATCTCGGCTGCCCCCTGCTGCAACCTCTCTTATTTCCAGATGGGACTA 8400
Db 8341 CTTCGATGAGCAATCTCGGCTGCCCCCTGCTGCAACCTCTCTTATTTCCAGATGGGACTA 8400
Oy 8401 CGCACTGCACTGGCCCTAAAGCTCACCACCAATCATCCAGATGGCTAGCCAGGAGACTCAG 8460
Db 8401 CGCACTGCACTGGCCCTAAAGCTCACCACCAATCATCCAGATGGCTAGCCAGGAGACTCAG 8460
Oy 8461 GGATATGCTGGGCTCTGCTCCACAGTGTATAAATACAGGATACATACGCTGGAGAG 8520
Db 8461 GGATATGCTGGGCTCTGCTCCACAGTGTATAAATACAGGATACATACGCTGGAGAG 8520
Oy 8521 ATTTTAACTGAATCCTGAGATAGAGCAGGCACTCTACCAATGGAGGTTCTTTTGT 8580
Db 8521 ATTTTAACTGAATCCTGAGATAGAGCAGGCACTCTACCAATGGAGGTTCTTTTGT 8580
Oy 8581 GTTTGGTTGGTTTCTCTGCTGCAATAGATCAGGAGCTGAAATAGTACCTGGGCTTAC 8640
Db 8581 GTTTGGTTGGTTTCTCTGCTGCAATAGATCAGGAGCTGAAATAGTACCTGGGCTTAC 8640
Oy 8641 ATTAACATCTTGTCTCAAAAGCCTATAGAGTAGGAGGCTGAGGCTTAAGAAGACCTT 8700
Db 8641 ATTAACATCTTGTCTCAAAAGCCTATAGAGTAGGAGGCTGAGGCTTAAGAAGACCTT 8700
Oy 8701 AAGCGGCTGTGATAGACACAGATAGCCTGCACTATATACCAAGACCTTGTTCAAAA 8760
Db 8701 AAGCGGCTGTGATAGACACAGATAGCCTGCACTATATACCAAGACCTTGTTCAAAA 8760
Oy 8761 ACATGGAGGAGGGGATGTTTAAAGTGTGGGCTGTGAACAGGACCTAAGGAGGCCAA 8820
Db 8761 ACATGGAGGAGGGGATGTTTAAAGTGTGGGCTGTGAACAGGACCTAAGGAGGCCAA 8820
Oy 8821 TGTAGACATTTGACTAAGAAGATCATCATCAAGCCGGTGGGAGGAGTGTGAGGTTTG 8880
Db 8821 TGTAGACATTTGACTAAGAAGATCATCATCAAGCCGGTGGGAGGAGTGTGAGGTTTG 8880
Oy 8881 ACTACAGTGTCAAGACCCCATAGAGACCGATTCTCTCTCTGAGGCTCAACG 8940
Db 8881 ACTACAGTGTCAAGACCCCATAGAGACCGATTCTCTCTCTGAGGCTCAACG 8940
Oy 8941 CTGGGCTCGAGGCGCACTGCTTCACATGCTCTCTCTAGGCTGCTCCACCATG 8995
Db 8941 CTGGGCTCGAGGCGCACTGCTTCACATGCTCTCTCTAGGCTGCTCCACCATG 8995

```

RESULT 2
 AAA49605
 ID AAA49605 standard: DNA; 8993 BP.

AC AAA49605;

DT 19-DEC-2000 (first entry)

DE Mouse villin gene regulatory elements and partial sequence.

KW Mouse; villin; intestinal epithelial cell;

XX uro-genital tract epithelial cell; tumour; ds.

OS Mus sp.

FT Key Location/Qualifiers

FT misc_signal 3442

FT /tag= a /note= "transcriptional start site"

FT exon 3442..3487

FT /tag= b /number= 1

```

FT intron 3488..8981
FT /tag= c
FT /number= 1
FT exon 8982..8995
FT /tag= d
FT /number= 2
FT misc_signal 8992
FT /tag= e /note= "translational start site"
PN WO20003493-A2.
XX 15-JUN-2000.
PD 09-DEC-1999; 99WO-EP09782.
PE 09-DEC-1998; 98WO-EP08009.
PR (CNRS ) CENT NAT RECH SCI.
PA (CURIE-) INST CURIE.
PI Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J;
DR WPI; 2000-423434/36.
XX Novel nucleotide sequence derived from mouse villin gene for targeted
XX expression of transgenes in immature and differentiated epithelial
XX cells of intestine or urogenital tracts -
PS Claim 3; Fig 6; 52pp; English.
CC The present sequence consists of the regulatory region and the first
CC exon and intron of the murine villin gene. This gene is expressed in the
CC epithelial cells of the intestine and uro-genital tracts. Its promoter
CC sequence can be used in the targeted expression of exogenous genes in
CC these places, which may, for example, be useful in the treatment of
XX tumours.
SQ Sequence 8993 BP; 2274 A; 2105 C; 2257 G; 2355 T; 2 other;
Query Match 96.2%; Score 8651; DB 21; Length 8993;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8991; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
Oy 1 GATCTGTGTCACCAAGAGACACTGTGTCACACTGGGAGGAGTGGAGAGAGGTCA 60
Db 1 GATCTGTGTCACCAAGAGACACTGTGTCACACTGGGAGGAGTGGAGAGAGGTCA 60
Oy 61 GAAGTTTAAAGTCAATCTGTGTTACATAGCAAGTTTCAGCCAGCTTCAGTACATGANA 120
Db 61 GAAGTTTAAAGTCAATCTGTGTTACATAGCAAGTTTCAGCCAGCTTCAGTACATGANA 120
Oy 121 CCTTGTGTTGTTGTTGTTGTTTAAAGCATTAATAATATCCATAAGAGAGTTGG 180
Db 121 CCTTGTGTTGTTGTTGTTGTTTAAAGCATTAATAATATCCATAAGAGAGTTGG 180
Oy 181 CAGTGTGTCAGACACCTTTAATCCAGTATTCAGAGAGCAGAAACAGGACAGATCTCTGT 240
Db 181 CAGTGTGTCAGACACCTTTAATCCAGTATTCAGAGAGCAGAAACAGGACAGATCTCTGT 240
Oy 241 GAGTTGCAAGTCAAGCTAGTCTGCAAGAGTACTCCAGAGTGGCAAGGGCTACACAGACA 300
Db 241 GAGTTGCAAGTCAAGCTAGTCTGCAAGAGTACTCCAGAGTGGCAAGGGCTACACAGACA 300
Oy 301 AACCTTGTCTCAATAAACCAAGTATGATAGTAAGCAATGAGAAATTTGGA 360
Db 301 AACCTTGTCTCAATAAACCAAGTATGATAGTAAGCAATGAGAAATTTGGA 360
Oy 361 GTTCATTCAAGATGAGACATCTATAGATGATCTCTTGACCCAGTAAAGTATGTGCA 420
Db 361 GTTCATTCAAGATGAGACATCTATAGATGATCTCTTGACCCAGTAAAGTATGTGCA 420
Oy 421 TGGGGAAGGAGGAGTGTCTAGATTAAAGTGTGAGGCGATGCTATTCTCA 480

```


| | | | |
|----|-------|---|--------|
| OY | 264.1 | AGAGAGGCTGCCTCTCAACACGATCTTCATCTTTCCCTCTTAAGAGACACGATTC | 2700 |
| Db | 264.1 | AGAAAGAGCTCCCTCTCAACACGATCTTCATCTTTCCCTCTTAAGAGAGACGATTC | 2700.0 |
| OY | 270.1 | AAGGGGAGAGATCTACAGGGGGGACAGGGGAGGGGAGGAGACAGCCATGGTTTCC | 2760 |
| Db | 270.1 | AAGGTGGCAGAAAGTTCTACAGGGGGGACAGGGGAGGGGAGGAGACAGCCATGGTTTCC | 2760.0 |
| OY | 276.1 | AGAGACTTACAGCAGAGGGGACGACAGGACATCCAGGTCCAGGGGAGGGAGGTGAGG | 2820 |
| Db | 276.1 | AGAGAGCTTACAGCAGAGGGGACGACAGGACATCCAGGTTCCAGGGGAGGGAGGTGAGG | 2820.0 |
| OY | 282.1 | CCCTTGTTCGAGGAGAGAGGACGGGGGCAAGACAGGGTTCAAAAGGCACAGTTTATGCA | 2880 |
| Db | 282.1 | CCCTTGTTCGAGGAGAGAGGACGGGGGCAAGACAGGGTTCAAAAGGCACAGTTTATGCA | 2880.0 |
| OY | 288.1 | GCTCATAAAGTGAGGTCCTGTGGCTCACTACAGAAAGAGAGAAAGGAGAGGCGCTTGT | 2940 |
| Db | 288.1 | GCTCATAAAGTGAGGTCCTGTGGCTCACTACAGAAAGAGAGAAAGGAGAGGCGCTTGT | 2940.0 |
| OY | 294.1 | GCCCACTGAGGAGAGGTCATGCTGAGTATAGAGAGATCTGACAGGGTGGCAGAGGCCAC | 3000 |
| Db | 294.1 | GCCCACTGAGGAGAGGTCATGCTGAGTATAGAGAGATCTGACAGGGTGGCAGAGGCCAC | 3000.0 |
| OY | 300.1 | CTGTCTGTCCCAAGGAGAACCCCAAGTGTAACCTGTGGCTTGGGTGCTGAGTTCAGCTA | 3060 |
| Db | 300.1 | CTGTCTGTCCCAAGGAGAACCCCAAGTGTAACCTGTGGCTTGGGTGCTGAGTTCAGCTA | 3060.0 |
| OY | 306.1 | CAGAGCCCAAGAGTCTCTACTCCATCCATCCAGTGGCCCCCTGGCCGACACCCCA | 3120 |
| Db | 306.1 | CAGAGCCCAAGAGTCTCTACTCCATCCATCCAGTGGCCCCCTGGCCGACACCCCA | 3120.0 |
| OY | 312.1 | CCCCGACCTCCCGGCCACTTCTCTAGGGCTGGAGGGTGGCCAGGCTGGTGGGGTTGC | 3180 |
| Db | 312.1 | CCCCGACCTCCCGGCCACTTCTCTAGGGCTGGAGGGTGGCCAGGCTGGTGGGGTTGC | 3180.0 |
| OY | 318.1 | CTACCTGACGATAGAGCCGACGTTCTTAGCCGGAAGTGACCCCATCTCTGAAGCTGAGA | 3240 |
| Db | 318.1 | CTACCTGACGATAGAGCCGACGTTCTTAGCCGGAAGTGACCCCATCTCTGAAGCTGAGA | 3240.0 |
| OY | 324.1 | GCCAAAGGGGGGGGACACAGGGGACACTGACGGCTGTGACGGCTGTGGGCTCTAGAGTTCC | 3300 |
| Db | 324.1 | GCCAAAGGGGGGGGACACAGGGGACACTGACGGCTGTGACGGCTGTGGGCTCTAGAGTTCC | 3300.0 |
| OY | 330.1 | AGGGAAGTGGGACACTCTATCTCCGACCCGCCCATCTCTCTGGGGGCCCTATCTTCC | 3360 |
| Db | 330.1 | AGGGAAGTGGGACACTCTATCTCCGACCCGCCCATCTCTCTGGGGGCCCTATCTTCC | 3360.0 |
| OY | 336.1 | CTTATATGTGTAAAGAGTTCCTGGGGGGGGGGGGTGGTGTGAGGACAAAGTCTTTCG | 3420 |
| Db | 336.1 | CTTATATGTGTAAAGAGTTCCTGGGGGGGGGGGGTGGTGTGAGGACAAAGTCTTTCG | 3420.0 |
| OY | 342.1 | GTTCTCTGACACCAGCTTGGCCACAACCTCTTAAGATCTCCAGGTGGTGGTGGCTCTTC | 3480 |
| Db | 342.1 | GTTCTCTGACACCAGCTTGGCCACAACCTCTTAAGATCTCCAGGTGGTGGTGGCTCTTC | 3480.0 |
| OY | 348.1 | CAGACAGTAAAGGCAATTTGGGTGGGACACATGTGACCCACAGGTGGTGGAGGGAGAG | 3540 |
| Db | 348.1 | CAGACAGTAAAGGCAATTTGGGTGGGACACATGTGACCCACAGGTGGTGGAGGGAGAG | 3540.0 |
| OY | 354.1 | GGTCCCTGCTTCTCTCTGGAGGCTGTGGCTTCTGAGACCTTGGTAAAGTTTGGGGG | 3600 |
| Db | 354.1 | GGTCCCTGCTTCTCTCTGGAGGCTGTGGCTTCTGAGACCTTGGTAAAGTTTGGGGG | 3600.0 |
| OY | 360.1 | TGAGTAAAGTGCTCTGAAACTCTGAAAGAGCAGAAACCCAGCAGGCTGTCTGGGCT | 3660 |
| Db | 360.1 | TGAGTAAAGTGCTCTGAAACTCTGAAAGAGCAGAAACCCAGCAGGCTGTCTGGGCT | 3660.0 |
| OY | 366.1 | TCATTAAGAGGATTTACACAGACCCCTTTCCTGTAAAGTACCTTGCTTCACTGTGTAG | 3720 |
| Db | 366.1 | TCATTAAGAGGATTTACACAGACCCCTTTCCTGTAAAGTACCTTGCTTCACTGTGTAG | 3720.0 |

| | | | |
|----|------|---|------|
| QY | 3721 | ATCCCTGGGACCAAGGTGGCTCCTGGGACTGACGATTTCTACAAATTTAAATCAGACAGT | 3780 |
| Db | 3721 | ATTCCTCTGGGACCAAGGTGGCTCCTGGGACTGACGATTTCTACAAATTTAAATCAGACAGT | 3780 |
| QY | 3781 | CCTGAGACTTGGACTCCGTCGTCGTATTTATCTACTCTCTCGGGGGCTCATTTCTGAGT | 3840 |
| Db | 3781 | CCTGAGACTTGGACTCCGTCGTCGTATTTATCTACTCTCTCGGGGGCTCATTTCTGAGT | 3840 |
| QY | 3841 | TCATGTCTTACACATCTGAAATGGTTTCTTTGTGTACCAATTCGCCCTGACACTCTCTGGGA | 3900 |
| Db | 3841 | TCATGTCTTACACATCTGAAATGGTTTCTTTGTGTACCAATTCGCCCTGACACTCTCTGGGA | 3900 |
| QY | 3901 | GGTGTATTCCTTGGCACATATCTTCGCGATGTATTAECTGACACACAGAGAGAGGGGG | 3960 |
| Db | 3901 | GGTGTATTCCTTGGCACATATCTTCGCGATGTATTAECTGACACACAGAGAGAGGGGG | 3960 |
| QY | 3961 | AGAGTCAGAGACTGTGTCTTAGGGCCATTAAGCCCTGGAGCAATCACCCCTTCTGTGAAT | 4020 |
| Db | 3961 | AGAGTCAGAGACTGTGTCTTAGGGCCATTAAGCCCTGGAGCAATCACCCCTTCTGTGAAT | 4020 |
| QY | 4021 | GGCCCTTCATTTTTTCGTTTACCATGATCTATTTTATACAGAGTGGGACGTGAAGACCA | 4080 |
| Db | 4021 | GGCCCTTCATTTTTTCGTTTACCATGATCTATTTTATACAGAGTGGGACGTGAAGACCA | 4080 |
| QY | 4081 | AACCTGCCAGAGTTTGGGACTCTACACAGCCAGGTTATCTGCTCAGAAATCCCTGT | 4140 |
| Db | 4081 | AACCTGCCAGAGTTTGGGACTCTACACAGCCAGGTTATCTGCTCAGAAATCCCTGT | 4140 |
| QY | 4141 | TCACCTTGAGGTGGGACAAATCTGCTCTGGGGGCTTCAGAGTCTTGTTAGCAGAGAGGT | 4200 |
| Db | 4141 | TCACCTTGAGGTGGGACAAATCTGCTCTGGGGGCTTCAGAGTCTTGTTAGCAGAGAGGT | 4200 |
| QY | 4201 | ATTCCTTTTATAGGGGCACTGACCTAGTCTATAGTGTTACTACAAATTCCTGTCCAGTTAAAG | 4260 |
| Db | 4201 | ATTCCTTTTATAGGGGCACTGACCTAGTCTATAGTGTTACTACAAATTCCTGTCCAGTTAAAG | 4260 |
| QY | 4261 | CTGGAACCTAAACCACGCGACGGCCAGGATTCCTACAGTTGTACCCCAAGAACAACA | 4320 |
| Db | 4261 | CTGGAACCTAAACCACGCGACGGCCAGGATTCCTACAGTTGTACCCCAAGAACAACA | 4320 |
| QY | 4321 | AGACGTAGATATCCAGAGATAGGTAGCTGGGGAGAAAGAACTTAAACCCCCCAAG | 4380 |
| Db | 4321 | AGACGTAGATATCCAGAGATAGGTAGCTGGGGAGAAAGAACTTAAACCCCCCAAG | 4380 |
| QY | 4381 | GGCCACAGTTCCTGTCCCTTAGTTTCAATATGCCAGTATAGGTACTACTATGGGCTG | 4440 |
| Db | 4381 | GGCCACAGTTCCTGTCCCTTAGTTTCAATATGCCAGTATAGGTACTACTATGGGCTG | 4440 |
| QY | 4441 | TGACTTGGTACTACAGCATAGATGATGTTCATGTGTGTAGTGTATATCTGAGCAC | 4500 |
| Db | 4441 | TGACTTGGTACTACAGCATAGATGATGTTCATGTGTGTAGTGTATATCTGAGCAC | 4500 |
| QY | 4501 | TTGGAGAGCTTAACACAGAGATTTGCTATATGTTTAGGGCCACGCTGAGACTATAGAGCA | 4560 |
| Db | 4501 | TTGGAGAGCTTAACACAGAGATTTGCTATATGTTTAGGGCCACGCTGAGACTATAGAGCA | 4560 |
| QY | 4561 | GACTTGTCTTTAGAAAAAAAATGAAGCCACACAGCTAGTGGGACACGGCTTTAATCCCA | 4620 |
| Db | 4561 | GACTTGTCTTTAGAAAAAAAATGAAGCCACACAGCTAGTGGGACACGGCTTTAATCCCA | 4620 |
| QY | 4621 | GCACCTTGGGAGGACGAAGCAGGCAATTTCTGAATTCAGAGGCGACGCTGTATAGAGT | 4680 |
| Db | 4621 | GCACCTTGGGAGGACGAAGCAGGCAATTTCTGAATTCAGAGGCGACGCTGTATAGAGT | 4680 |
| QY | 4681 | GAGTTCCAGGACACGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAACAGAAAAACA | 4740 |
| Db | 4681 | GAGTTCCAGGACACGCCAGGGCTACACAGAGAAACCTGTTTGAAGAAACAGAAAAACA | 4740 |
| QY | 4741 | AACCAAAACAAAACAAAACAAAACCCAAACCCAAACCCAAACCTCTCATCTCTCATCTCTC | 4800 |
| Db | 4741 | AACCAAAACAAAACAAAACAAAACCCAAACCCAAACCCAAACCTCTCATCTCTCATCTCTC | 4800 |
| QY | 4801 | TAGGCTGTGTCTGTCTAGTGAGTATGAGTTTGGGGACTTGTAGACTTATATATATATAGGCC | 4860 |

|||||
Db 4801 TAGGCTGTCTCTAGTGTAGAGTTTGGGACTTCAGACTATATATTAATAGGCC 4860
QY 4861 TTTTATATCATCTGTGAGAGACAGAAAGTTTCAGTCTGTGGGACACAGTGGGACCTGAGAA 4920
Db 4861 TTTTATATCATCTGTGAGAGACAGAAAGTTTTCAGTCTGTGGGACACAGTGGGACCTGAGAA 4920
QY 4921 AAGTACTCTCTCCAGCCAAAATTTCTGGGAAGCTTCTGTGGGAAGTGTGTCCGAT 4980
Db 4921 AAGTACTCTCTCCAGCCAAAATTTCTGTGGGAAGCTTCTGTGGGAAGTGTGTCCGAT 4980
QY 4981 CAGACTACTGTCTAGAAAGCAGAAAGAGAGGGTTGGAAGAAATGTTGCTGGACAGACAGTT 5040
Db 4981 CAGACTACTGTCTAGAAAGCAGAAAGAGAGGGTTGGAAGAAATGTTGCTGGACAGACAGTT 5040
QY 5041 GGAACAGAAAGCAGAGGGGGGAGGCATCCAGATTTGCAATATGTAGCTAGCTTTTGGT 5100
Db 5041 GGAACAGAAAGCAGAGGGGGGAGGCATCCAGATTTGCAATATGTAGCTAGCTTTTGGT 5100
QY 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
Db 5101 TCTGTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
QY 5161 AATGATTCAGTTGAGGAGACATCCAGCCAGGGTCTTGTGGCAGCTTAAAGATG 5220
Db 5161 AATGATTCAGTTGAGGAGACATCCAGCCAGGGTCTTGTGGCAGCTTAAAGATG 5220
QY 5221 AAGGCCCTAACCCCTCCGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCCTCTA 5280
Db 5221 AAGGCCCTTAACCCCTCCGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCCTCTA 5280
QY 5281 GGGTGAAGAGAGATATCTGTCTGTACCAACATGGCTAGAGACAGAAAGCATTTGACACAG 5340
Db 5281 GGGTGAAGAGAGATATCTGTCTGTACCAACATGGCTAGAGACAGAAAGCATTTGACACAG 5340
QY 5341 TTAACCCCTCAGAACCAAGCCATCCCTTGGCTTAAGAGAGCTGGGCCCTTTTCGTTT 5400
Db 5341 TTAACCCCTCAGAACCAAGCCATCCCTTGGCTTAAGAGAGCTGGGCCCTTTTCGTTT 5400
QY 5401 AAGAACTTACTTTTCTTCAGAGAGAGCAAGGCTTTGGCCCTCCCTGTGGTCAA 5460
Db 5401 AAGAACTTACTTTTCTTCAGAGAGAGCAAGGCTTTGGCCCTCCCTGTGGTCAA 5460
QY 5461 TAAACACCCCTGTGTATACATTAATTTTACTGTGCTAGTTGCTCCAGAGACAGTCA 5520
Db 5461 TAAACACCCCTGTGTATACATTAATTTTACTGTGCTAGTTGCTCCAGAGACAGTCA 5520
QY 5521 TCTGTAGACCTGTCTCTAATCAACAGGTATGGCCACATTCCTCACCCAGAGAG 5580
Db 5521 TCTGTAGACCTGTCTCTAATCAACAGGTATGGCCACATTCCTCACCCAGAGAG 5580
QY 5581 TGCACAAAGAGACCTTAGAGAAAGGTATACATTAACAAAGATGGCCAGATTAACAAAA 5640
Db 5581 TGCACAAAGAGACCTTAGAGAAAGGTATACATTAACAAAGATGGCCAGATTAACAAAA 5640
QY 5641 ACTACTATCCCTTGTATACCAAAATTTGCTGTGAACACAGAGGGGTGTGAGTGTAT 5700
Db 5641 ACTACTATCCCTTGTATACCAAAATTTGCTGTGAACACAGAGGGGTGTGAGTGTAT 5700
QY 5701 GT 5760
Db 5701 GT 5760
QY 5761 CTGTGGGGGACTTTTATGCTAAGAAATATCTATATTTGGCCCATGGCCAAAGGGGTAT 5820
Db 5761 CTGTGGGGGACTTTTATGCTAAGAAATATCTATATTTGGCCCATGGCCAAAGGGGTAT 5820
QY 5821 TGGGAGAGACTCAGGCTTGTGCAAAACAGTAAGCTGCCAAGATGGATTGTGGCTGAA 5880
Db 5821 TGGGAGAGACTCAGGCTTGTGCAAAACAGTAAGCTGCCAAGATGGATTGTGGCTGAA 5880
QY 5881 TACCAAGAGGGGAGGCTGATCAGAGTGCAGACATCAAGATTAAGCCACCCTGTGGG 5940
|||||

Db 5881 TCACCAAGGGGAGGCTGATCAGAGTGCAGACATCAAGATTAAGCCACCCTGTGGG 5940
QY 5941 GCTCAGAAAGGAGTATTTTCAAGAGGTAAAGGCCAAAGCATTTATTTATCCAGACATGAC 6000
Db 5941 GCTCAGAAAGGAGTATTTTCAAGAGGTAAAGGCCAAAGCATTTATTTATCCAGACATGAC 6000
QY 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGTGAGAGATGGGCTGTAGTGTGGGACAC 6060
Db 6001 TCAAAATCAAAAGTGCAGAGAGATTTAGCTGTGAGAGATGGGCTGTAGTGTGGGACAC 6060
QY 6061 TGAACCTTGACATTTATAGTCTAGAGCCAGAGAGATGCACAGAGGTGATGSGTCTTA 6120
Db 6061 TGAACCTTGACATTTATAGTCTAGAGCCAAAGAGAGATGCACAGAGGTGATGSGTCTTA 6120
QY 6121 CTACAGCTTGAGACAGGACAGTGGAGAAATGGGACTCCATCCATGATGAGAGAGGCTGAG 6180
Db 6121 CTACAGCTTGAGACAGGACAGTGGAGAAATGGGACTCCATCCATGATGAGAGAGGCTGAG 6180
QY 6181 CACCAACAGGTACAAGTGTCCCTGTGTCTCATGCCAGAGATTCCTGGCCAGTTTCAAG 6240
Db 6181 CACCAACAGGTACAAGTGTCCCTGTGTCTCATGCCAGAGATTCCTGGCCAGTTTCAAG 6240
QY 6241 GACTAAGGACTCATCTGTGTGTGAAAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300
Db 6241 GACTAAGGACTCATCTGTGTGTGAAAACAAAGTATCCAGCCCTAAGCCCATTTTGTCT 6300
QY 6301 AATTTAATCAGAAACCCCTGGGATGAGGCTGAGGACAGACAGAGCTTTTAAAAAGCTC 6360
Db 6301 AATTTAATCAGAAACCCCTGGGATGAGGCTGAGGACAGACAGAGCTTTTAAAAAGCTC 6360
QY 6361 CCAGGTATTTGATCAGAGCTGAGAACAAACACAGCTCAGGTTTCAAAACAGAAAGAGC 6420
Db 6361 CCAGGTATTTGATCAGAGCTGAGAACAAACACAGCTCAGGTTTCAAAACAGAAAGAGC 6420
QY 6421 AAAGCTAAGGAAAGCTTGGGATGGGAGGCTTCTTCAGGCCAGTAGATGAGGCTGGTT 6480
Db 6421 AAAGCTAAGGAAAGCTTGGGATGGGAGGCTTCTTCAGGCCAGTAGATGAGGCTGGTT 6480
QY 6481 AAGCAGGTGGGAGGCTTCTGTGCTGTGATATAGCTATCCATCCAGTCAATCCAT 6540
Db 6481 AAGCAGGTGGGAGGCTTCTGTGCTGTGATATAGCTATCCATCCAGTCAATCCAT 6540
QY 6541 ACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCCATCCATCCATCCAT 6600
Db 6541 ACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCCATCCATCCATCCAT 6600
QY 6601 CCAGCATCCATCCAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCAGCATCCATCCAAACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
QY 6661 CATTTATCCAAAGAGAACTGATTTGATTAAGTGGGAGATTAATTAATTTTGA 6720
Db 6661 CATTTATCCAAAGAGAACTGATTTGATTAAGTGGGAGATTAATTAATTTTGA 6720
QY 6721 ACCTGTGTGATTTGACTGATTTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 6780
Db 6721 ACCTGTGTGATTTGACTGATTTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 6780
QY 6781 TGGCAATCGGAGAAAGGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840
Db 6781 TGGCAATCGGAGAAAGGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840
QY 6841 TTGAACCTCAAAATTTATCGGCTGTGGCAAGTCTTTTACACCGAGCCATTTTGTCTGACA 6900
Db 6841 TTGAACCTCAAAATTTATCGGCTGTGGCAAGTCTTTTACACCGAGCCATTTTGTCTGACA 6900
QY 6901 CATCATTTATTTAGAAAGCATCTTATGTATGTCCAGGCTGAGGCTCAAGCTTCTATGTGCG 6960
Db 6901 CATCATTTATTTAGAAAGCATCTTATGTATGTCCAGGCTGAGGCTCAAGCTTCTATGTGCG 6960
QY 6961 CCAGGATGACCTTTTAACTCTGCTCTTCAGGCTTCACCGAGTGTAGTTTACAGGT 7020
Db 6961 CCAGGATGACCTTTTAACTCTGCTCTTCAGGCTTCACCGAGTGTAGTTTACAGGT 7020
6960 CCAGGATGACCTTTTAACTCTGCTCTTCAGGCTTCACCGAGTGTAGTTTACAGGT 7019

| | | | |
|----|------|--|------|
| QY | 7021 | GTTCACACTGGTGAATGECCTTTAATCCAGCACTCTGTGGGGGGGGGGGGGAGCGGATC | 7080 |
| Db | 7020 | GTTCACACTGGTGAATGECCTTTAATCCAGCACTCTGTGGGGGGGGGGGGGAGCGGATC | 7079 |
| QY | 7081 | CTGTGGTGGAGGGCCACTTTTGGTCTCAGAGGTTTCAGATTAACCTGGGGCTATACAGGAA | 7140 |
| Db | 7080 | CCTGAGTGGAGGGCCACTTTTGGTCTCAGAGGTTTCAGATTAACCTGGGGCTATACAGGAA | 7139 |
| QY | 7141 | ACCCATATCCCAACAAACAACAACAAAAAATTTCTGTGCATTAATCACAGAGA | 7200 |
| Db | 7140 | ACCCATATCCCAACAAACAACAACAACAAAAAATTTCTGTGCATTAATCACAGAGA | 7199 |
| QY | 7201 | TTAGAGGATATTACTAGGGTAGTAGGGCGGTGAGGAGAGCAATGCTTTCTTTGTAT | 7260 |
| Db | 7200 | TTAGAGGATATTACTAGGGTAGTAGGGCGGTGAGGAGAGCAATGCTTTCTTTGTAT | 7259 |
| QY | 7261 | ATAATAGTAAGATACCTCAAGAAGATGCATTAATCATATCATATCATATCATATCATATC | 7320 |
| Db | 7260 | ATAATAGTAAGATACCTCAAGAAGATGCATTAATCATATCATATCATATCATATCATATC | 7319 |
| QY | 7321 | TATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATC | 7380 |
| Db | 7320 | TATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATCATATC | 7379 |
| QY | 7381 | CTGCTTTGACCTCAAAAGCTCTATTCTTGGGGCAACTCTTACCCTAGTGTGGGTTT | 7440 |
| Db | 7380 | CTGCTTTGACCTCAAAAGCTCTATTCTTGGGGCAACTCTTACCCTAGTGTGGGTTT | 7439 |
| QY | 7441 | ACCAACACCCAGACATTTATTTTATTTTATTTTATTTATTAATCTAGAGCTCAGGGTG | 7500 |
| Db | 7440 | ACCAACACCCAGACATTTATTTTATTTTATTTTATTTATTTATTAATCTAGAGCTCAGGGTG | 7499 |
| QY | 7501 | GGACACAGAGGCTGTGTCATGCTATAGCAAGCTCTGCCACAGACTGAGCTCCAGTCC | 7560 |
| Db | 7500 | GGACACAGAGGCTGTGTCATGCTATAGCAAGCTCTGCCACAGACTGAGCTCCAGTCC | 7559 |
| QY | 7561 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTTCATATTCGCAGCGCTATGATGCTCTC | 7620 |
| Db | 7560 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTTCATATTCGCAGCGCTATGATGCTCTC | 7619 |
| QY | 7621 | CACGCCCAAGTCTCAGAGCACTTGTGGCAATCCAGTGGGGGGGCAACCTGTGGCTACCA | 7680 |
| Db | 7620 | CACGCCCAAGTCTCAGAGCACTTGTGGCAATCCAGTGGGGGGGCAACCTGTGGCTACCA | 7679 |
| QY | 7681 | GTGCCCTGTTCCTCCTGTCTCAGACACTACATATTTGCTGTGCAAGCTTCAATGTAATG | 7740 |
| Db | 7680 | GTGCCCTGTTCCTCCTGTCTCAGACACTACATATTTGCTGTGCAAGCTTCAATGTAATG | 7739 |
| QY | 7741 | GGATGCGTTCCTGTGATTTCTTTTATGCGTGGGCCCTTATCTTAGCACAGTTGTGTG | 7800 |
| Db | 7740 | GGATGCGTTCCTGTGATTTCTTTTATGCGTGGGCCCTTATCTTAGCACAGTTGTGTG | 7799 |
| QY | 7801 | GGCCATGTGCTACGCTATCACTCATCTATCATCATCTTATAGGCTTAATAGTGTCTT | 7860 |
| Db | 7800 | GGCCATGTGCTACGCTATCACTCATCTATCATCATCTTATAGGCTTAATAGTGTCTT | 7859 |
| QY | 7861 | TGTGTGATTAACACACTTCTGTTTCACTTACTGATGAATTTGTGGCCCAACCCCCAC | 7920 |
| Db | 7860 | TGTGTGATTAACACACTTCTGTTTCACTTACTGATGAATTTGTGGCCCAACCCCCAC | 7919 |
| QY | 7921 | CCTTTTATTTTATTTTATTTGACAAAGGCTTGTGTGTATCTTGGCAATCTGGCGTCT | 7980 |
| Db | 7920 | CCTTTTATTTTATTTTATTTGACAAAGGCTTGTGTGTATCTTGGCAATCTGGCGTCT | 7979 |
| QY | 7981 | GAGCTCACTCTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGTAGACACTCTGTGAC | 8040 |
| Db | 7980 | GAGCTCACTCTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGTAGACACTCTGTGAC | 8039 |
| QY | 8041 | AGAGTACCATGAACCTTCAAAAGCAATTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTG | 8100 |
| Db | 8040 | AGAGTACCATGAACCTTCAAAAGCAAA - TTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTG | 8098 |

| | | | |
|----|------|---|------|
| OY | 8101 | TGTAATGCTGTAATATGTCATGTTTGTGCTTCACAGTGTCTGCAATGCTGTACCTGTGTG | 8160 |
| Db | 8099 | TGTATGCGCTGTATATGTGCATGTTTGTGTCTTCAGAGTGTCTCACATGTGTATACCTGTGTG | 8158 |
| OY | 8161 | TGGACACAGAGAACAAACCGATGTGCATTCCTTCAGATATACAGCATCTTGTAAATATGTA | 8220 |
| Db | 8159 | TGGGACAGAGAACAAACCGATGTGTGCATTCCTTCAGATATACAGCATCTTGTAAATATGTA | 8218 |
| OY | 8221 | TGTATATATATGTTTATTTTAGTGTGCCAAGTATGCAAGTATTTTGTGTGGAGTTTTCAC | 8280 |
| Db | 8219 | TGTATATATATGTTTATTTTAGTGTGCCAAGTATGCAAGTATTTTGTGTGGAGTTTTCAC | 8278 |
| OY | 8281 | CTTCCCTGTGTGGGCTCTCCGCATTTAAATCAGTCTCTGGGCTAGTAGAGCAATGGCTTCA | 8340 |
| Db | 8279 | CTTCCCTGTGTGGGCTCTCCGCATTTAAATCAGTCTCTGGGCTAGTAGAGCAATGGCTTCA | 8338 |
| OY | 8341 | CTCGATGAGCCATCTGCGTCCCTGCTGCACTCTCTTATTTTCCAGATGGGACTA | 8400 |
| Db | 8339 | CTCGATGAGCCATCTGCGTCCCTGCTGCACTCTCTTATTTTCCAGATGGGACTA | 8398 |
| OY | 8401 | CGCACTGCACTGGGCTTAACACTACCAATGTCATCCAGATGGCTTAGCCAGGAGACTCAG | 8460 |
| Db | 8399 | CGCACTGCACTGGGCTTAACACTACCAAGTCATCCAGATGGCTTAGCCAGGAGACTCAG | 8458 |
| OY | 8461 | GGATATGCTGGGCTCTGCGCTCCACAGTGTGCAGTAATACAGGATACATCACTGCTGGAG | 8520 |
| Db | 8459 | GGATATGCTGGGCTCTGCGCTCCACAGTGTGCAGTAATACAGGATACATCACTGCTGGAG | 8518 |
| OY | 8521 | ATTTTAACTGTAATCTCTAGAGATAGACAGGCACTCTACCAATGGAGGTTCTTTTGT | 8580 |
| Db | 8519 | ATTTTAACTGTAATCTCTAGAGATAGACAGGCACTCTACCAATGGAGGTTCTTTTGT | 8578 |
| OY | 8581 | GTTTGGTTGGTTCCCTGTCATATAGATCAGGAGTCTCAATATAGTAGGCTGGGCTAC | 8640 |
| Db | 8579 | GTTTGGTTGGTTCCCTGTCATATAGATCAGGAGTCTCAATATAGTAGGCTGGGCTAC | 8638 |
| OY | 8641 | ATAACATCTTGTCTCAAAAAGCCTATAGAGGTAGGAGGTCGAGGCTAAGAAGAGCCT | 8700 |
| Db | 8639 | ATAACATCTTGTCTCAAAAAGCCTATAGAGGTAGGAGGTCGAGGCTAAGAAGAGCCT | 8698 |
| OY | 8701 | AAGCGGCTGTGATAGCACACAGATAGGCTGACATATATAGCAAGACCTTGTTCACAA | 8760 |
| Db | 8699 | AAGCGGCTGTGATAGCACACAGATAGGCTGACATATATAGCAAGACCTTGTTCACAA | 8758 |
| OY | 8761 | ACATGAGGAGGAGGTATGTTTTAGTGTGTGGGCTGTGTAAACAGGCACTAAGAGGAGCAA | 8820 |
| Db | 8759 | ACATGAGGAGGAGGTATGTTTTAGTGTGTGGGCTGTGTAAACAGGCACTAAGAGGAGCAA | 8818 |
| OY | 8821 | TGTATGACATTTGACTAAGAAAGATATATCATCAAGCCGGGTGGGCAAGGTGAAGTTGG | 8880 |
| Db | 8819 | TGTATGACATTTGACTAAGAAAGATATATCAAAACCGGGGTGGGCAAGGTGAAGTTGG | 8878 |
| OY | 8881 | ACTACAGTGTGTCAAGACCCCATAGAGAACCAATTTCCCTTTCTCTGGGGCTCAAGC | 8940 |
| Db | 8879 | ACTACAGTGTGTCAAGACCCCATAGAGAACCAATTTCCCTTTCTCTGGGGCTCAAGC | 8938 |
| OY | 8941 | CTGAGCTGACAGGCCACTGTCTCTACATGCCTTCTCTCTAGGCTGTGCCACATG | 8995 |
| Db | 8939 | CTGAGCTGACAGGCCACTGTCTCTACATGCCTTCTCTCTAGGCTGTGCCACATG | 8993 |

RESULT 3
AAS19044
ID AAS19044 standard; cDNA; 2628 BP.
XX
XX
AC AAS19044;
XX
DT 09-Apr-2002 (first entry)
XX
XX
DE Mouse MU-1 haematopoietin receptor superfamily chain cDNA sequence.
DE
XX
XX
KW Mouse; MU-1; haematopoietin receptor superfamily chain; thyroiditis;
RV immune deficiency; anaemia; autoimmune disorder; multiple sclerosis;
FN


```
XX 22-OCT-2001 (first entry)
XX
XX
DE Human polynucleotide SEQ ID NO 5076.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0596042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Mehran T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB: AAM41931.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 5076; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 2447 BP; 733 A; 387 C; 515 G; 812 T; 0 other;
XX
XX Query Match 0.7%; Score 64; DB 22; Length 2447;
XX Best Local Similarity 100.0%; Pred. No. 8.6e-17;
XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 6
AAS13715/C
ID AAS13715 standard; DNA; 80 BP.
XX
XX AAS13715;
XX
XX 08-MAY-2002 (first entry)
XX
XX Simple sequence repeat, SSR, #12.
XX
XX Simple sequence repeat; plant; ds; SSR; ryegrass; fescue; tandem repeat;
XX cereal profiling; grass profiling; seed batch purity testing.
XX
XX Poeae.
XX
XX NZ509193-A.
XX
XX 25-MAY-2001.
XX
XX 03-JAN-2001; 2001NZ-0509193.
XX
XX 24-DEC-1999; 99AU-0004906.
XX
XX 04-MAY-2000; 2000AU-0007310.
XX
XX (SAUS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
XX (USC-) UNIV SOUTHERN CROSS.
XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
XX (UVAD-) UNIV ADELAIDE.
XX (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX
XX Forster JW, Jones ES;
XX
XX WPI: 2001-512563/56.
XX
XX New simple sequence repeats having 2 or more tandemly repeated
XX nucleotide core elements isolated from ryegrass and fescue, useful for
XX selecting of genes in grass or cereal breeding or profiling grass or
XX cereal species varieties -
XX
XX Claim 6, Page 51; 72pp; English.
XX
XX The invention relates to a substantially purified or isolated nucleic
XX acid (1) from ryegrass or fescue species including a simple sequence
XX repeat (SSR), having 2 or more tandemly repeated nucleotide core elements
XX 2-6 nucleotides in length. Also included are a nucleic acid primer
XX suitable for amplifying an SSR, identifying (M1) an SSR by preparing a
XX library of ryegrass or fescue genomic DNA enriched for SSRs and
XX identifying clones in the library containing SSRs, a library of ryegrass
XX or fescue genomic DNA enriched for SSRs prepared by the M1, selecting for
XX a gene in grass or cereal breeding by identifying an SSR that is closely
XX associated with the gene such that the SSR and the gene are
XX preferentially co-inherited, and selecting for the SSR in the
XX breeding, a method for DNA profiling grass or cereal species varieties by
XX assessing variation between SSR varieties and testing the purity of grass
XX or cereal seed batches by assessing variation within seed batch of an
XX SSR. The SSRs may be used in the selection of genes in grass or cereal
XX breeding, for profiling grass or cereal species varieties, for testing
XX the purity of grass or cereal seed batches, and for DNA profiling to
XX establish the distinct identity, uniformity and/or stability of a
XX cultivar. The present sequence is a ryegrass or fescue SSR.
XX
XX Sequence 80 BP; 40 A; 33 C; 0 G; 7 T; 0 other;
XX
XX Query Match 0.7%; Score 63; DB 23; Length 80;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-16;
XX Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


PT diagnosis of cell proliferative disorders, e.g. cancer, AIDS, moles,
PT warts and psoriasis.
XX
PS Claim 4; Page 29-31a; 40pp; English.
XX
CC 3 Human chromosome-4 fragments (AAT32033-35) comprise genetic
CC determinants that are capable of inhibiting or arresting the
CC proliferative capacity of a cell. They were identified using a
CC microcell-mediated chromosome transfer protocol. The fragments
CC are located at 4q26-28 or 4q33-34.1. They can be used in gene
CC therapy of cancer and diseases, such as AIDS, that involve
CC proliferative cells. They can also be used to diagnose the extent
CC and/or severity of malignancy.
XX
SQ Sequence 4580 BP; 1218 A; 1067 C; 1052 G; 1243 T; 0 other;
XX
Query Match 0.7%; Score 63; DB 17; Length 4580;
Best Local Similarity 100.0%; Pred. No. 2.2e-16;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2234 TGGTGGCTCACAACCATCATGATCTGACACCCCTCTTGTGGTGCATCGAAGACA 2293
DB 3141 TGGTGGCTCACAACCATCATGATCTGACACCCCTCTTGTGGTGCATCGAAGACA 3082
OY 2294 GCT 2296
DB 3081 GCT 3079
XX
RESULT 12
AAV83940/C
ID AAV83940 standard; DNA; 80240 BP.
XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KM Yeast artificial chromosome: YAC; probe: eukaryotic chromosome;
KM neocentromere; replication; extra-chromosomal element; segregation;
KM cell division; artificial chromosome; gene therapy; mardel(10);
KM human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PE 13-MAY-1998; 98MO-AU00352.
XX
PR 26-AUG-1997; 97AU-0008791.
PR 13-MAY-1997; 97AU-0006784.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Cancellia MR, Choo K, Du Sart D;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains
CC an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This

CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating; acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins, and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 8 other;
XX
Query Match 0.7%; Score 63; DB 20; Length 80240;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5700 TG 5759
DB 52119 TG 52060
OY 5760 TCT 5762
DB 52059 TCT 52057
XX
RESULT 13
AAQ27094
ID AAQ27094 standard; DNA; 219 BP.
XX
AC AAQ27094;
XX
DT 25-JAN-1993 (first entry)
XX
DE Fragile X diagnostic primer 3.
XX
KM Fragile X locus; (CCG)_n repeat; Xq27.3; PCR; CpG island;
KM polymerase chain reaction; AC repeat; microsatellite; ss.
XX
OS Homo sapiens.
XX
PN WO9212262-A.
XX
PD 23-JUL-1992.
XX
PE 03-JAN-1992; 92MO-US00020.
XX
PR 04-JAN-1991; 91US-0638518.
PR 20-MAR-1991; 91US-0672232.
PR 06-DEC-1991; 91US-0802650.
XX
PA (ADEL-) ADELATDE MEDICAL CENT WOMEN & CHILDREN.
PA (UNITW) UNITV WASHINGTON.
XX
PI Nagaraja R, Richards RI, Schlessinger D, Sutherland GR;
XX
DR WPI; 1992-268680/32.
XX
PT DNA spanning the Fragile X site on the human X chromosome - for
PT diagnosis and treatment of associated mental retardation
XX
PS Claim 15; Page 27; 47pp; English.
XX
CC The sequences given in AAQ27092-100 are oligonucleotide primers which
CC correspond to sequences flanking polymorphic microsatellite AC repeat
CC sequences within the human Fragile X locus. These primers may be used
CC to identify alleles of the microsatellite regions which vary in AC
CC repeat length. This allows a method for screening for a microsatellite

Thu Jan '30 09:11:34 2003

us-09-877-935-1.rng

Page 17

Db 1605 GT 1604

Search completed: January 29, 2003, 14:09:44
Job time : 1956 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2003, 11:46:17 : Search time 21787 Seconds
(without alignments)
12015.410 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctgtgtcaccaagaca.....ctctagctcgtcaccaatg 8995

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pal:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|------------|--------------------|
| 1 | 8995 | 100.0 | 8995 | 6 | AX027824 | AX027824 Sequence |
| 2 | 4465 | 49.6 | 192060 | 2 | AC098570 | AC098570 Mus muscu |
| 3 | 96 | 1.1 | 149109 | 2 | AC101018 | AC101018 Rattus no |
| 4 | 96 | 1.1 | 159492 | 2 | AC115173 | AC115173 Rattus no |
| 5 | 82 | 0.9 | 222185 | 2 | AC127697 | AC127697 Mus muscu |
| 6 | 82 | 0.9 | 226909 | 2 | AC079490 | AC079490 Mus muscu |
| 7 | 82 | 0.9 | 239077 | 2 | AC093351 | AC093351 Mus muscu |
| 8 | 80 | 0.9 | 235302 | 2 | AC073784 | AC073784 Mus muscu |
| 9 | 79 | 0.9 | 202855 | 10 | AL589737 | AL589737 Mouse DNA |
| 10 | 78 | 0.9 | 204068 | 2 | AC103395 | AC103395 Mus muscu |
| 11 | 77 | 0.9 | 174982 | 2 | AC110034 | AC110034 Mus muscu |
| 12 | 77 | 0.9 | 184854 | 2 | AC122433 | AC122433 Mus muscu |
| 13 | 76 | 0.8 | 178032 | 2 | AC087869 | AC087869 Mus muscu |
| 14 | 76 | 0.8 | 201602 | 2 | AC074146 | AC074146 Mus muscu |
| 15 | 76 | 0.8 | 203083 | 2 | AC069459 | AC069459 Mus muscu |
| 16 | 76 | 0.8 | 245958 | 2 | AL731687 | AL731687 Mus muscu |
| 17 | 75 | 0.8 | 148180 | 2 | AC113443 | AC113443 Mus muscu |
| 18 | 75 | 0.8 | 168889 | 10 | AL731779 | AL731779 Mouse DNA |
| 19 | 75 | 0.8 | 191080 | 2 | AC083914 | AC083914 Mus muscu |
| 20 | 75 | 0.8 | 200370 | 2 | AC112940 | AC112940 Mus muscu |
| 21 | 75 | 0.8 | 263350 | 2 | AC130210 | AC130210 Mus muscu |
| 22 | 74 | 0.8 | 178428 | 2 | AC108601 | AC108601 Rattus no |
| 23 | 74 | 0.8 | 221378 | 2 | AC122835 | AC122835 Mus muscu |
| 24 | 74 | 0.8 | 271723 | 2 | AC122210 | AC122210 Mus muscu |
| 25 | 73 | 0.8 | 15000 | 10 | AF420002 | AF420002 Mus muscu |
| 26 | 73 | 0.8 | 110000 | 2 | AC125151_2 | Continuation (3 of |
| 27 | 73 | 0.8 | 187271 | 2 | AC102745 | AC102745 Mus muscu |
| 28 | 73 | 0.8 | 192086 | 2 | AC117634 | AC117634 Mus muscu |
| 29 | 73 | 0.8 | 205336 | 10 | AC116576 | AC116576 Mus muscu |
| 30 | 73 | 0.8 | 222973 | 2 | AC093403 | AC093403 Mus muscu |
| 31 | 73 | 0.8 | 238186 | 2 | AC112676 | AC112676 Mus muscu |
| 32 | 71 | 0.8 | 106605 | 9 | AL353691 | AL353691 Human DNA |
| 33 | 71 | 0.8 | 110716 | 2 | AC097770 | AC097770 Rattus no |
| 34 | 71 | 0.8 | 153727 | 10 | AC093449 | AC093449 Mus muscu |
| 35 | 71 | 0.8 | 166509 | 2 | AC117101 | AC117101 Rattus no |
| 36 | 71 | 0.8 | 191895 | 2 | AL844872 | AL844872 Mus muscu |
| 37 | 71 | 0.8 | 197632 | 10 | AL672039 | AL672039 Mouse DNA |
| 38 | 71 | 0.8 | 214000 | 10 | AC093317 | AC093317 Mus muscu |
| 39 | 70 | 0.8 | 149930 | 2 | AC130543 | AC130543 Mus muscu |
| 40 | 70 | 0.8 | 150552 | 2 | AC121952 | AC121952 Mus muscu |
| 41 | 70 | 0.8 | 176814 | 10 | AL627444 | AL627444 Mouse DNA |
| 42 | 70 | 0.8 | 184167 | 2 | AC118211 | AC118211 Mus muscu |
| 43 | 70 | 0.8 | 205615 | 2 | AC117826 | AC117826 Mus muscu |
| 44 | 70 | 0.8 | 211096 | 2 | AC127309 | AC127309 Mus muscu |
| 45 | 70 | 0.8 | 211348 | 10 | AC098883 | AC098883 Mus muscu |

ALIGNMENTS

RESULT 1
AX027824 LOCUS AX027824 8995 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent WO0034492.
ACCESSION AX027824
VERSION AX027824.1 GI:10188668
KEYWORDS
SOURCE
ORGANISM Mus sp.
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 8995)
AUTHORS Robine,S., Louvard,D., Pinto,D. and Jaisser,F.
TITLE Regulatory sequences of the mouse villin gene - use in transgenesis
JOURNAL Patent: WO 0034492-A 1 15-JUN-2000;

Db 1921 AGGACTCATATGGTGGCTCACAGCCATCTGTAAATCCAGTTCCAGAGGGTTCCACACCC 1980
Qy 1981 CTTCTGGCCTCCACAGGCACACATATAGTACACAGACATACATGCGAGGCAAAACACC 2040
Db 1981 CTTCTGGCCTCCACAGGCACACATATAGTACACAGACATACATGCGAGGCAAAACACC 2040
Qy 2041 CATACACATATAATTAATAGGAACTTAAAGGTGTCATGTGTTGTTAAACATTTGTCT 2100
Db 2041 CATACACATATAATTAATAGGAACTTAAAGGTGTCATGTGTTGTTAAACATTTGTCT 2100
Qy 2101 TACACATGCTGATGTAAGACATGTACACGCACACACTGAAGGGATCTGGGGCTGGAG 2160
Db 2101 TACACATGCTGATGTAAGACATGTACACGCACACACTGAAGGGATCTGGGGCTGGAG 2160
Qy 2161 AGATGGCTAGGGGTTAAGAGCAGCTAGCTCTCTCCGAAGGAAGGTCCGATTTCAAT 2220
Db 2161 AGATGGCTAGGGGTTAAGAGCAGCTAGCTCTCTCCGAAGGAAGGTCCGATTTCAAT 2220
Qy 2221 CCTAGCACCCATAGTGGTGCACAAACCATCCATATGAGATGACACCCCTTCTGTG 2280
Db 2221 CCTAGCACCCATAGTGGTGCACAAACCATCCATATGAGATGACACCCCTTCTGTG 2280
Qy 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATCTAATAATAATCTTTT 2340
Db 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATCTAATAATAATCTTTT 2340
Qy 2341 TTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGATTAATAGCAGTACACG 2400
Db 2341 TTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGATTAATAGCAGTACACG 2400
Qy 2401 GGTGATTAATCTATCTGAGATTTTTCCTTCCGCTTGGCTTGCACAGTGGTGACAGAG 2460
Db 2401 GGTGATTAATCTATCTGAGATTTTTCCTTCCGCTTGGCTTGCACAGTGGTGACAGAG 2460
Qy 2461 CCCCTTTTCAATTACAGAAGAGGGGTCTACATTAATTTCTGACAAAAACAGCAGCTCAG 2520
Db 2461 CCCCTTTTCAATTACAGAAGAGGGGTCTACATTAATTTCTGACAAAAACAGCAGCTCAG 2520
Qy 2521 ATGTACTGCTCTGCTGCTGATGAGCAGGGCAGGGGCGGCGGACACACACACACAC 2580
Db 2521 ATGTACTGCTCTGCTGCTGATGAGCAGGGCAGGGGCGGCGGACACACACACACAC 2580
Qy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA 2640
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA 2640
Qy 2641 AGAAGAGGCTGCCCTCAAAACGATCTTATCTTCCCTCTAAGAGACACGATTC 2700
Db 2641 AGAAGAGGCTGCCCTCAAAACGATCTTATCTTCCCTCTAAGAGACACGATTC 2700
Qy 2701 AAGGTGGCAGAAGATCTACAGGGGGGAGAGGGGAGAGGAGGCAATGATTTCC 2760
Db 2701 AAGGTGGCAGAAGATCTACAGGGGGGAGAGGGGAGAGGAGGCAATGATTTCC 2760
Qy 2761 AGAGACCTACAGCAGAGGGCAGCAAGAGATCCAGGTCCAGGGCAGGGAGAGTGGAG 2820
Db 2761 AGAGACCTACAGCAGAGGGCAGCAAGAGATCCAGGTCCAGGGCAGGGAGAGTGGAG 2820
Qy 2821 CCCTTTCTCCGAGAGAGGAGGCGGAGACAGAGGTTCAAGGCAAGGTTATGGCA 2880
Db 2821 CCCTTTCTCCGAGAGAGGAGGCGGAGAACAGGGTTCAAGGCAAGGTTATGGCA 2880
Qy 2881 GGTCAATAAAGTGGAGGTGGTGCATCAGAAAGAGAGAAAGGAAAGGCCCTTGT 2940
Db 2881 GGTCAATAAAGTGGAGGTGGTGCATCAGAAAGAGAGAAAGGAAAGGCCCTTGT 2940
Qy 2941 GCCCAGTGAAGAGGATCTGAGTGAAGAGATCTGAGGGGTGCCAGAGGCCCA 3000
Db 2941 GCCCAGTGAAGAGGATCTGAGTGAAGAGATCTGAGGGGTGCCAGAGGCCCA 3000
Qy 3001 CTGTCTGTCCAAAGGAAACCCCAAGTGTGAATCTGTGGCTGTGATTTCCAGCTA 3060
Db 3001 CTGTCTGTCCAAAGGAAACCCCAAGTGTGAATCTGTGGCTGTGATTTCCAGCTA 3060

Qy 3061 CAAGACCCCAAGAGTCTACTCCATCCATCCAGTCCAGTCCGCCCCGCCCCACACCCA 3120
Db 3061 CAAGACCCCAAGAGTCTACTCCATCCATCCAGTCCAGTCCGCCCCGCCCCACACCCA 3120
Qy 3121 CCCCCAGTCCCGTGCACATCTCTAGGGCTGAGAGGTGGCCAGCCCTGTGGGGTTGC 3180
Db 3121 CCCCCAGTCCCGTGCACATCTCTAGGGCTGAGAGGTGGCCAGCCCTGTGGGGTTGC 3180
Qy 3181 CTACTGCAAGTGAAGACCCAGTCTAGCCGGAAGTGCACCCCATCTGAACTGCAGA 3240
Db 3181 CTACTGCAAGTGAAGACCCAGTCTAGCCGGAAGTGCACCCCATCTGAACTGCAGA 3240
Qy 3241 GCCAAGGGGGGGCACAACGGCAGCTAGGCTGACAGGCTGTGCTGAGTGGCTTGTCC 3300
Db 3241 GCCAAGGGGGGGCACAACGGCAGCTAGGCTGACAGGCTGTGCTGAGTGGCTTGTCC 3300
Qy 3301 AGGACCTGGGGCAGCTACTTCCACCCCCCATCCATCTCTGAGGGCCATCTTCC 3360
Db 3301 AGGACCTGGGGCAGCTACTTCCACCCCCCATCCATCTCTGAGGGCCATCTTCC 3360
Qy 3361 CTTATATGCTGAAGAAATTCTGGGGGGGGGGTGTGTGAGACAAAGTCTTGC 3420
Db 3361 CTTATATGCTGAAGAAATTCTGGGGGGGGGGTGTGTGAGACAAAGTCTTGC 3420
Qy 3421 GTCTCTGCAAGCCAGCTTGGCACAACCTTCCTAAGATCTCCAGTGGTGGCTCTTC 3480
Db 3421 GTCTCTGCAAGCCAGCTTGGCACAACCTTCCTAAGATCTCCAGTGGTGGCTCTTC 3480
Qy 3481 CAGACAGTGAAGCAATTTGGTGGGACACATGTATGACACAGTGGTTGAGAGGAGAC 3540
Db 3481 CAGACAGTGAAGCAATTTGGTGGGACACATGTATGACACAGTGGTTGAGAGGAGAC 3540
Qy 3541 GGTCTTCTCTCTCTGTGGAGCTGTGCTTGTGTGACACTTGGTATTAAGTTGGGG 3600
Db 3541 GGTCTTCTCTCTCTGTGGAGCTGTGCTTGTGTGACACTTGGTATTAAGTTGGGG 3600
Qy 3601 TGAAGTGAAGTGTCTGAAGCTGAAAGCTGAAAGAGCAAGAGCAGAGCTGTGGGCT 3660
Db 3601 TGAAGTGAAGTGTCTGAAGCTGAAAGCTGAAAGAGCAAGAGCAGAGCTGTGGGCT 3660
Qy 3661 TCAATGAAGAAATTGACAGAGCCCTTTCCTGTAAGTACACTTCCTCATCTGTAG 3720
Db 3661 TCAATGAAGAAATTGACAGAGCCCTTTCCTGTAAGTACACTTCCTCATCTGTAG 3720
Qy 3721 ATTCCCTGGAGCAAGGTGCTCTGGGACTCAGATTTCTACATTTAAATCAGGACGT 3780
Db 3721 ATTCCCTGGAGCAAGGTGCTCTGGGACTCAGATTTCTACATTTAAATCAGGACGT 3780
Qy 3781 CCTGAGACTTGGACGTCGCGCTGTATTTACTACTCTCTGCTGGCTGATTTCTGTG 3840
Db 3781 CCTGAGACTTGGACGTCGCGCTGTATTTACTACTCTCTGCTGGCTGATTTCTGTG 3840
Qy 3841 TCATGTCTTACACATCTGAATGTTCTTGTGTGTCACCATTCCTGCACTCTGGGA 3900
Db 3841 TCATGTCTTACACATCTGAATGTTCTTGTGTGTCACCATTCCTGCACTCTGGGA 3900
Qy 3901 GGTGATCTCTTGGCAGATGTATCTGGGATGTAACTGACAGCCACAGAGAGAGGGG 3960
Db 3901 GGTGATCTCTTGGCAGATGTATCTGGGATGTAACTGACAGCCACAGAGAGAGGGG 3960
Qy 3961 AGAGTAGAGAGTGTCTGAGGCTTATTAAGGCTGAGACATACCCCTTCTAGAAAT 4020
Db 3961 AGAGTAGAGAGTGTCTGAGGCTTATTAAGGCTGAGACATACCCCTTCTAGAAAT 4020
Qy 4021 GGGCCCTCCATTTTGGTATCCATGATCTATTTATACAGAGTGGGAGTGAAGGCA 4080
Db 4021 GGGCCCTCCATTTTGGTATCCATGATCTATTTATACAGAGTGGGAGTGAAGGCA 4080
Qy 4081 AACCTGCCAAGAAATTGGGACTCACTCAGACCAGGTTATCTGCTCAGAAATCCCTTG 4140
Db 4081 AACCTGCCAAGAAATTGGGACTCACTCAGACCAGGTTATCTGCTCAGAAATCCCTTG 4140

| | | | |
|----|-------|---|-------|
| OY | 414.1 | TCACCTGAGGTTGGGAAATCTCCCTCTGGGGCCTCCAGGCTTGTTAGAGAGAGGT | 4 200 |
| Db | 414.1 | TCACCTGAGGTTGGGAAATCTCCCTCTGGGGCCTCCAGGCTTGTTAGAGAGAGGT | 4 200 |
| OY | 420.1 | ATCCTTGTATAGGGCATGACACTAGTCTATAGGTTACTACATATTCCTGCTCCAGTTAAAG | 4 260 |
| Db | 420.1 | ATCCTTGTATAGGGCATGACACTAGTCTATAGGTTACTACATATTCCTGCTCCAGTTAAAG | 4 260 |
| OY | 426.1 | CTGGAACCTAAAACCCACGAGCGGCCAGGATTTCTCTACAGTTGTACCCAGAACAACA | 4 320 |
| Db | 426.1 | CTGGAACCTAAAACCCACGAGCGGCCAGGATTTCTCTACAGTTGTACCCAGAACAACA | 4 320 |
| OY | 432.1 | AGACAGTAGATATACAGAGATAGTAGTCTGGGAGAAAGAACTTTAAACCCCCCAAG | 4 380 |
| Db | 432.1 | AGACAGTAGATATACAGAGATAGTAGTCTGGGAGAAAGAACTTTAAACCCCCCAAG | 4 380 |
| OY | 438.1 | GCCCAAGAGTCCGTTCCCTAGTTCCAAATGCCAGATAGTAGTACAGTATAGGGGTG | 4 440 |
| Db | 438.1 | GCCCAAGAGTCCGTTCCCTAGTTCCAAATGCCAGATAGTAGTACAGTATAGGGGTG | 4 440 |
| OY | 444.1 | TGAGTTGGTACTACAAGCATGATGATGTTCTATGTTAGTGTATATATCTAGCAC | 4 500 |
| Db | 444.1 | TGAGTTGGTACTACAAGCATGATGATGTTCTATGTTAGTGTATATATCTAGCAC | 4 500 |
| OY | 450.1 | TTTGGAGGCTGAAAGCAGAGAGATTGCTATATGTGTTGAGGCCACCTGAGCTATAGACGA | 4 560 |
| Db | 450.1 | TTTGGAGGCTGAAAGCAGAGAGATTGCTATATGTGTTGAGGCCACCTGAGCTATAGACGA | 4 560 |
| OY | 456.1 | GACTTGTCTTTAAGAAAAAATGAAGAAACCCACAGTGTGTGACACGCTTTAAATCCCA | 4 620 |
| Db | 456.1 | GACTTGTCTTTAAGAAAAAATGAAGAAACCCACAGTGTGTGTGACACGCTTTAAATCCCA | 4 620 |
| OY | 462.1 | GCACCTGGGAGGACGAAGCAGCATTTTCTGAGTTCAAGGGCAGCCTGTTATAGAGT | 4 680 |
| Db | 462.1 | GCACCTGGGAGGACGAAGCAGCATTTTCTGAGTTCAAGGGCAGCCTGTTATAGAGT | 4 680 |
| OY | 468.1 | GAGTTCACGACACCCAGGCTACACAGAGAAACCTGTTTGAAGAAACACAGAAAAACA | 4 740 |
| Db | 468.1 | GAGTTCACGACACCCAGGCTACACAGAGAAACCTGTTTGAAGAAACACAGAAAAACA | 4 740 |
| OY | 474.1 | AACAAACAAACAAACAAACCAACCCAAACCCAAACCCAAACCTCTCATCTCTC | 4 800 |
| Db | 474.1 | AACAAACAAACAAACAAACCAACCCAAACCCAAACCCAAACCTCTCATCTCTC | 4 800 |
| OY | 480.1 | TAGGCTGTCTGTCTAGTGTGTAGATTTGGGGACTTGAGACTTATATATAAATAGGC | 4 860 |
| Db | 480.1 | TAGGCTGTCTGTCTAGTGTGTAGATTTGGGGACTTGAGACTTATATATAAATAGGC | 4 860 |
| OY | 486.1 | TTTTTATTCACCTGGTCAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTGTAGA | 4 920 |
| Db | 486.1 | TTTTTATTCACCTGGTCAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTGTAGA | 4 920 |
| OY | 492.1 | AAGTACCTCTGGCAGCCCAAAATTTCTGGGAAAGCTTCTGTGAGGAAGTGTGCCAT | 4 980 |
| Db | 492.1 | AAGTACCTCTGGCAGCCCAAAATTTCTGGGAAAGCTTCTGTGAGGAAGTGTGCCAT | 4 980 |
| OY | 498.1 | CAGACTACTGTTTACAGAGCAGAGAGAGGTTTGAAGAAATGTTGTGAGACAGACTT | 5 040 |
| Db | 498.1 | CAGACTACTGTTTACAGAGCAGAGAGAGGTTTGAAGAAATGTTGTGAGACAGACTT | 5 040 |
| OY | 504.1 | GGAACAGAGAGACAGAGGGGGAGGATCCAGATTTCTGACATGTAGCTGACTTTTGGT | 5 100 |
| Db | 504.1 | GGAACAGAGAGACAGAGGGGGAGGATCCAGATTTCTGACATGTAGCTGACTTTTGGT | 5 100 |
| OY | 510.1 | TCTCTGGGTGACAAGTGTCCCCCAGGGATATAGGCTGTGAAGAAAGGGGACACGAGGTGACCC | 5 160 |
| Db | 510.1 | TCTCTGGGTGACAAGTGTCCCCCAGGGATATAGGCTGTGAAGAAAGGGGACACGAGGTGACCC | 5 160 |
| OY | 516.1 | AATGAGTTCAAGTTGAGAGGACATCAGGCCAGGGTCTTCTGTGGCAAGCTAAAGAGT | 5 220 |
| Db | 516.1 | AATGAGTTCAAGTTGAGAGGACATCAGGCCAGGGTCTTCTGTGGCAAGCTAAAGAGT | 5 220 |
| OY | 522.1 | AGAGCCCTTAACCTTCCCTGAAGTTTAGGGGAGACAGAGAGCTGAGAGATCCTCTTA | 5 280 |

| | | | | |
|----|------|--|--|------|
| Db | 5221 | AGACCCCTCPAACCCCTCCCGAAGTTTACGGGGAGACGAGAGGCTGAGAGAAATCCCTTCTA | | 5280 |
| OY | 5281 | GGGTGAAGAGAGAGTATCTCTCTGTGACCAACAATGCGCTAGAGACAGAAACAGTTGGACAG | | 5340 |
| Db | 5281 | GGGTGAAGAGAGAGTATCTCTCTGTGACCAACAATGCGCTAGAGACAGAAACAGTTGGACAG | | 5340 |
| OY | 5341 | TTTACCCCTCAGAACACGACATCCCTCTTGGCTCTAAGAGGCTGGGCCCTTCTCTGTT | | 5400 |
| Db | 5341 | TTTACCCCTCAGAACACGACATCCCTCTTGGCTCTAAGAGGCTGGGCCCTTCTCTGTT | | 5400 |
| OY | 5401 | AAGAAATCTTACTTTTCTTCAGAGAGGACGACAGCTTTGTCCCCCTCCCTGTGTCTCA | | 5460 |
| Db | 5401 | AAGAAATCTTACTTTTCTTCAGAGAGGACGACAGCTTTGTCCCCCTCCCTGTGTCTCA | | 5460 |
| OY | 5461 | TAAACACCCCGTGTGTGAATATAGTTTATTTTACTGTGATGTTGCTCTCAGACAGTCCA | | 5520 |
| Db | 5461 | TAAACACCCCGTGTGTGAATATAGTTTATTTTACTGTGATGTTGCTCTCAGACAGTCCA | | 5520 |
| OY | 5521 | TCTGTGTAGACCTGTGCTCTTACTCTACCAAGATATGGCCACATTCTCTACCCAGAAAG | | 5580 |
| Db | 5521 | TCTGTGTAGACCTGTGCTCTTACTCTACCAAGATATGGCCACATTCTCTACCCAGAAAG | | 5580 |
| OY | 5581 | TGCGAAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAAGATGGCCAGCAATTAACAACA | | 5640 |
| Db | 5581 | TGCGAAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAAGATGGCCAGCAATTAACAACA | | 5640 |
| OY | 5641 | ACTACTATCCCTTTGTACCCAAATTTGGTTTTGCGAACCGAGGAGGGGTGTGTGAGTGTAT | | 5700 |
| Db | 5641 | ACTACTATCCCTTTGTACCCAAATTTGGTTTTGCGAACCGAGGAGGGGTGTGTGAGTGTAT | | 5700 |
| OY | 5701 | GTTGT | | 5760 |
| Db | 5701 | GTTGT | | 5760 |
| OY | 5761 | CTTGGGGACTTTTCATGCTAAAGAAATATCTGATATTTGGGCCCTTCGCAACAGGGGTAT | | 5820 |
| Db | 5761 | CTTGGGGACTTTTCATGCTAAAGAAATATCTGATATTTGGGCCCTTCGCAACAGGGGTAT | | 5820 |
| OY | 5821 | TGGGGAGGTGAGGCTTCTGCAAAACAGTAAGCTGCCCAACATGAGATTTGGGGCTTAA | | 5880 |
| Db | 5821 | TGGGGAGGTGAGGCTTCTGCAAAACAGTAAGCTGCCCAACATGAGATTTGGGGCTTAA | | 5880 |
| OY | 5881 | TCACCAAGGGGACAGGCTGATCAGAGTGTGACAGAACATCAACAAGATAAGCCACCCCTGTGG | | 5940 |
| Db | 5881 | TCACCAAGGGGACAGGCTGATCAGAGTGTGACAGAACATCAACAAGATAAGCCACCCCTGTGG | | 5940 |
| OY | 5941 | GCTCAGAAGAGGAGTTTACAGAGGTTAAAGGCCAAGCCATTATTTATTCACAAGACATGAC | | 6000 |
| Db | 5941 | GCTCAGAAGAGGAGTTTACAMAGGTTAAAGGCCAAGCCATTATTTATTCACAAGACATGAC | | 6000 |
| OY | 6001 | TCAAAATCAAAAGTGCAAGGAGATATAGCTGGAGAGATGGGGCTTCAGTGTGGGACACC | | 6060 |
| Db | 6001 | TCAAAATCAAAAGTGCAAGGAGATATAGCTGGAGAGATGGGGCTTCAGTGTGGGACACC | | 6060 |
| OY | 6061 | TGACCTTGCACCTTATTATAGTACTAGGCCCAAGAGCAGTTCACAGAGGTTGATCTGGTCTCTA | | 6120 |
| Db | 6061 | TGACCTTGCACCTTATTATAGTACTAGGCCCAAGAGCAGTTCACAGAGGTTGATCTGGTCTCTA | | 6120 |
| OY | 6121 | CTCAGCTTTGGAGCAGGACGTGAGAGATGAGTCCATCTCTGATGAGAGGGCTGAG | | 6180 |
| Db | 6121 | CTCAGCTTTGGAGCAGGACGTGAGAGATGAGTCCATCTCTGATGAGAGGGCTGAG | | 6180 |
| OY | 6181 | CACCAACAGGTACCAAGTGTCCCTGTGTCTCATGTGCCAGGATTCGCGCAATTTTCAAG | | 6240 |
| Db | 6181 | CACCAACAGGTACCAAGTGTCCCTGTGTCTCATGTGCCAGGATTCGCGCAATTTTCAAG | | 6240 |
| OY | 6241 | GACTAAGAGACTCATCTCTGTGTGAAACCAAGTATCCAGCCCTTAAGCCCATTTTGGTCT | | 6300 |
| Db | 6241 | GACTAAGAGACTCATCTCTGTGTGAAACCAAGTATCCAGCCCTTAAGCCCATTTTGGTCT | | 6300 |
| OY | 6301 | AATTAATTCAGAACCCCTGGGGATTCGAGGCTCTGACAGCAGAGACTTTTAAAAAGCTC | | 6360 |

Db 6301 AATTAAATCAGAAACCCCTGGGATGCGAGGCTCTGAGCAGCAGGAGCTTTTAAAGAGCTC 6360
OY 6361 CCAGGTGATTTCTGATCAGGAGCTGGAACAACACAGCTACAGGTTCAAGACAGAAAGAGC 6420
Db 6361 CCAGGTGATTTCTGATCAGGAGCTGGAACAACACAGCTACAGGTTCAAGACAGAAAGAGC 6420
OY 6421 AAAGCTAGGAAAAGCTTGGGATG6GGAGGCGCTTCCAGGCCAGTAGATGAGAGCTGTG 6480
Db 6421 AAAGCTAGGAAAAGCTTGGGATG6GGAGGCGCTTCCAGGCCAGTAGATGAGAGCTGTG 6480
OY 6481 AGCAGTGGTGGAGCTTCTCTGCGCTGTCAATATAGCTATCCATCCAGCTCATCCAT 6540
Db 6481 AGCAGTGGTGGAGCTTCTCTGCGCTGTCAATATAGCTATCCATCCAGCTCATCCAT 6540
OY 6541 ACACCCAGCCATCCATTTATGACCCATCCCTTCATCCATCCATCCATCCATCCATCCAC 6600
Db 6541 ACACCCAGCCATCCATTTATGACCCATCCCTTCATCCATCCATCCATCCATCCATCCAC 6600
OY 6601 CCACGATCCATCCAAACCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCACGATCCATCCAAACCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
OY 6661 CATTTATCCACAGAGAGACTGATTTGATTAATGTGGAGATTTAATTAATTTTGA 6720
Db 6661 CATTTATCCACAGAGAGACTGATTTGATTAATGTGGAGATTTAATTAATTTTGA 6720
OY 6721 AGCTGCTGATTTGATGATTTGATGATGATGAGAGATGATGATGATGATGATGATGATG 6780
Db 6721 AGCTGCTGATTTGATGATTTGATGATGATGAGAGATGATGATGATGATGATGATGATG 6780
OY 6781 TGGCAATCGGAGAAAGTTTGGGTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
Db 6781 TGGCAATCGGAGAAAGTTTGGGTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
OY 6841 TTGAACTCAAATTTATGAGGCTGGTGGCAAGTCTTTACCAGCCAGCCATTTTGTGACA 6900
Db 6841 TTGAACTCAAATTTATGAGGCTGGTGGCAAGTCTTTACCAGCCAGCCATTTTGTGACA 6900
OY 6901 CATCATTTATTTAGAAAGATTTATGATGATGAGGCTGGGCTCAAGCTTGTATGTCG 6960
Db 6901 CATCATTTATTTAGAAAGATTTATGATGATGAGGCTGGGCTCAAGCTTGTATGTCG 6960
OY 6961 CCACGATGACCTTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
Db 6961 CCACGATGACCTTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
OY 7021 GTTCAACTGCTGATGCTTTTAACT 7080
Db 7021 GTTCAACTGCTGATGCTTTTAACT 7080
OY 7081 CCGAGTTGGAGGCGAGTTTGTCTACAGATTTCAAGATACCTGGGCTATTCAGGGAA 7140
Db 7081 CCGAGTTGGAGGCGAGTTTGTCTACAGATTTCAAGATACCTGGGCTATTCAGGGAA 7140
OY 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7200
Db 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 7200
OY 7201 TTAGAGAGATTTAGTGGGATGATGAGGCTGGTGGAGAGATGATGATGATGATGATGATG 7260
Db 7201 TTAGAGAGATTTAGTGGGATGATGAGGCTGGTGGAGAGATGATGATGATGATGATGATG 7260
OY 7261 AATAATGTAAGATCTACAAGATGATTAATCTATCTATCTATCTATCTATCTATCTATCTATC 7320
Db 7261 AATAATGTAAGATCTACAAGATGATTAATCTATCTATCTATCTATCTATCTATCTATCTATC 7320
OY 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380
OY 7381 CTGCTTTGACTGAGATGCTCTCTATTTCTGGGTCACACTCTCTACCCCTAGTGTGGGTTT 7440
Db 7381 CTGCTTTGACTGAGATGCTCTCTATTTCTGGGTCACACTCTCTACCCCTAGTGTGGGTTT 7440

OY 7441 ACCAACCACAGACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500
Db 7441 ACCAACCACAGACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500
OY 7501 GGACTGAGGCTTGTGATGCTTAAGCAGCTCTGACACAGAGCTGACGCTCAGCTC 7560
Db 7501 GGACTGAGGCTTGTGATGCTTAAGCAGCTCTGACACAGAGCTGACGCTCAGCTCAGCTC 7560
OY 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTTGTATTTTGGACGCTATGTAGCTCTCTC 7620
Db 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTTGTATTTTGGACGCTATGTAGCTCTCTC 7620
OY 7621 CACCTCCAGTTCCAGCAGCTTCTGCTATCCAGAGGGGGGAGCACTGCTGCTACCA 7680
Db 7621 CACCTCCAGTTCCAGCAGCTTCTGCTATCCAGAGGGGGGAGCACTGCTGCTACCA 7680
OY 7681 GTGCCCTGTTCCTGTCTGACAGCTACATATTTTGGCTGTGAGAGATTTCAATGTAATG 7740
Db 7681 GTGCCCTGTTCCTGTCTGACAGCTACATATTTTGGCTGTGAGAGATTTCAATGTAATG 7740
OY 7741 GATGCTTCTCTGTATTTCTTTATGCTGGCCCTTTATCTTACACAGTTTGTGTTG 7800
Db 7741 GATGCTTCTCTGTATTTCTTTATGCTGGCCCTTTATCTTACACAGTTTGTGTTG 7800
OY 7801 GGCATGCTGACCTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860
Db 7801 GGCATGCTGACCTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860
OY 7861 TGTGTGATTAACCACTTCTGTCTATTTACAGATGAATTTTGGGCCACCCCCAC 7920
Db 7861 TGTGTGATTAACCACTTCTGTCTATTTACAGATGAATTTTGGGCCACCCCCAC 7920
OY 7921 CCTTTTTTTTTTATTTAGACAGAGCTTCTGTGTATCTTGCATCTTGCTCTCT 7980
Db 7921 CCTTTTTTTTTTATTTAGACAGAGCTTCTGTGTATCTTGCATCTTGCTCTCTCT 7980
OY 7981 GAGCTACTCTGTAGACAGAGGCTGTGAGGCTGCTCTCACTTTGACACTCTGTTGAC 8040
Db 7981 GAGCTACTCTGTAGACAGAGGCTGTGAGGCTGCTCTCACTTTGACACTCTGTTGAC 8040
OY 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 8100
Db 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 8100
OY 8101 TGTATGCTGTATATGTGATGTTTGTCTTCAAGTGTCTCACATGTGTGTACCTGTGTG 8160
Db 8101 TGTATGCTGTATATGTGATGTTTGTCTTCAAGTGTCTCACATGTGTGTACCTGTGTG 8160
OY 8161 TGGGACAGAGAAACCAACCGATGTGCCATCTCAGATPACTAGGCATCTGTTAATATGTA 8220
Db 8161 TGGGACAGAGAAACCAACCGATGTGCCATCTCAGATPACTAGGCATCTGTTAATATGTA 8220
OY 8221 TGTATATGATGTTTATTTAGTGTGCCCAAGTATCAGGATTTTGTGAGATTTTAC 8280
Db 8221 TGTATATGATGTTTATTTAGTGTGCCCAAGTATCAGGATTTTGTGAGATTTTAC 8280
OY 8281 CTTCCTTGTGGCTCTCCGATTTAACTACAGCTCTCGGGCTAGTGAAGAGCTTTCA 8340
Db 8281 CTTCCTTGTGGCTCTCCGATTTAACTACAGCTCTCGGGCTAGTGAAGAGCTTTCA 8340
OY 8341 CTGAGTGAAGCATCTGCGAGCCCTGCTGACACTCTCTCTTTATTTTCCAGATGGAGCTA 8400
Db 8341 CTGAGTGAAGCATCTGCGAGCCCTGCTGACACTCTCTCTTTATTTTCCAGATGGAGCTA 8400
OY 8401 CGCAGTCACTGGCTTAAAGCTTACCAAGTCAATCAGAGTGGCTAGCCAGGAGACTCAG 8460
Db 8401 CGCAGTCACTGGCTTAAAGCTTACCAAGTCAATCAGAGTGGCTAGCCAGGAGACTCAG 8460
OY 8461 GATATGCTGGGCTCTGCTCCACAGTGTGAATTTACAGGCTATACATCACTGCTGGAG 8520
Db 8461 GATATGCTGGGCTCTGCTCCACAGTGTGAATTTACAGGCTATACATCACTGCTGGAG 8520

| | | | |
|----|------|--|------|
| QY | 8521 | ATTTTTAACTGATCCTGAGATTAAGCAGGACCTCACAAATGAGAGGTTCTTTTGT | 8580 |
| Db | 8521 | ATTTTAACTGAACTCTGAGGATTAAGCAGGACCTCACAAATGAGAGGTTCTTTTGT | 8580 |
| QY | 8581 | GTTTGGTTTGTTCTCTCTGCATTAAGATCAGGACGCTGGAATTAAGTGAAGCTTGAGCTAC | 8640 |
| Db | 8581 | GTTTGGTTTGTTGTTCTCTCTGCATTAAGATCAGGACGCTGGAATTAAGTGAAGCTTGAGCTAC | 8640 |
| QY | 8641 | ATAACATCTTGCTCCAAAAGCCTATAGAGGTAAAGGAGTGCAGAGCTAAACAGACCTT | 8700 |
| Db | 8641 | ATAACATCTTGCTCCAAAAGCCTATAGAGGTAAAGGAGTGCAGAGCTAAACAGAGCCTT | 8700 |
| QY | 8701 | AAGCGGCTGTGATAGCACACAGATAGGCTGCACATATATAGACAGACCTGTTTCAAA | 8760 |
| Db | 8701 | AAGCGGCTGTGATAGCACACAGATAGGCTGCACATATATAGACAGACCTGTTTCAAA | 8760 |
| QY | 8761 | NCATGAGGAGGAGGTATGTTTTAACTGCTGGGCTGTGTACAGGACCTAAAGAGCCAA | 8820 |
| Db | 8761 | NCATGAGGAGGAGGTATGTTTTAACTGCTGGGCTGTGTACAGGACCTAAAGAGCCAA | 8820 |
| QY | 8821 | TGTAGACATTGTGACTAGAAGAGATCATCAAAACCGGCTGGGACAGGATAGAGTTGG | 8880 |
| Db | 8821 | TGTAGACATTGTGACTAGAAGAGATCATCAAAACCGGCTGGGACAGGATAGAGTTGG | 8880 |
| QY | 8881 | ACTACAGTGTCTCAAGACCCCATAGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC | 8940 |
| Db | 8881 | ACTACAGTGTCTCAAGACCCCATAGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC | 8940 |
| QY | 8941 | CTGGCTCGAGGGCCACGTGCTCTACACATGGCTTCTCCTTAGGCTCGTCCACCATG | 8995 |
| Db | 8941 | CTGGCTCGAGGGCCACGTGCTCTACACATGGCTTCTCCTTAGGCTCGTCCACCATG | 8995 |

RESULT 2
 AC098570
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

AC098570 192060 bp DNA linear HTG 20-AUG-2002
 Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered
 pieces.
 AC098570
 AC098570.2 GI:22325297
 HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_FULFILLTOP.
 house mouse.
 Mus musculus
 Eumariota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 192060)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-278N11
 unpublished
 2 (bases 1 to 192060)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Bouckigalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, J.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodde, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gorf, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
 Lamezares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McKernan, P., McKernan, K., McPheters, R., Meltrim, J.,
 Menusz, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, K., Subramanian, A., Talamas, J., Teejaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zemdeh, L., Zimmer, A. and Zody, M.
 Direct Submission

JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
SOURCE

Submitted (24-Oct-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 192060)

Barren,B., Nusbbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Birtin,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalov,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A., Cook,A., Cooke,P., DeArliano,K., Dewar,K., Diaz J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gaidyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamal,A., Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Maylor,J., Matthews,C., McCarthy,M., Meldrum,J., Menais,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Noddu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodor,B., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Ziemek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (20-Aug-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 20, 2002 this sequence version replaced gi:16356898.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L14796
Center clone name: 278.N.11

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190464 bases at least Q40
Consensus quality: 191061 bases at least Q30
Consensus quality: 191345 bases at least Q20
Insert size: 163000; agarose-tp
Insert size: 191560; sum-of-ctgigs
Quality coverage: 12.4 in Q20 bases; agarose-tp
Quality coverage: 10.5 in Q20 bases; sum-of-ctgigs

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 4921: contig of 4921 bp in length
4922 5021: gap of 100 bp
5022 15200: contig of 10179 bp in length
15201 15300: gap of 100 bp
15301 78692: contig of 63392 bp in length
78693 78792: gap of 100 bp
78793 116118: contig of 37326 bp in length
116119 116219: gap of 100 bp
116219 174491: contig of 58273 bp in length
174492 174591: gap of 100 bp
174592 192060: contig of 17469 bp in length.

Location/Qualifiers

1. 192060
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-278N11"
/clone_lib="RP21-73 Female Mouse BAC"

```
misc_feature 1..4921
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
misc_feature 5022..15200
                /note="assembly_fragment"
misc_feature 15301..78692
                /note="assembly_fragment"
misc_feature 78793..116118
                /note="assembly_fragment"
misc_feature 116219..174491
                /note="assembly_fragment"
misc_feature 174592..192060
                /note="assembly_fragment"
                clone_end:r7
BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others
ORIGIN
Query Match 49.6% Score 4465; DB 2: Length 192060;
Best Local Similarity 99.9% Pred. No. 0:
Matches 4975: Conservative 0; Mismatches 3; Indels 4; Gaps 2:
Oy 1 GATCTGCTGACCAAGACACTGTGCTCCAGCACTGGGAGGTGGAGGAGAGGCTCA 60
Db 40161 GATCTGCTGACCAAGACACTGTGCTCCAGCACTGGGAGGTGGAGGAGAGGCTCA 40220
Oy 61 GAAGTTAAGCATCCTTGTTTACATACCAAGTTTCCAGCTTCAGCTACATGA 120
Db 40221 GAAGTTAAGCATCCTTGTTTACATACCAAGTTTCCAGCTTCAGCTACATGA 40280
Oy 121 CCTTGTGTTGTTGTTGTTTAAAGCATTAATAATACCATTAAGGAGTTGG 180
Db 40281 CCTTGTGTTGTTGTTGTTTAAAGCATTAATAATACCATTAAGGAGTTGG 40340
Oy 181 CAGTGTGTCAGACACCTTAAATCCAGTATTCAGAGGCGAAGACGAGATCTGT 240
Db 40341 CAGTGTGTCAGACACCTTAAATCCAGTATTCAGAGGCGAAGACGAGATCTGT 40400
Oy 241 GAGTTCGAAGTACGCTACTGCAAGAGTATTCAGAGTGGCAAGGGCTACACAG 300
Db 40401 GAGTTCGAAGTACGCTACTGCAAGAGTATTCAGAGTGGCAAGGGCTACACAG 40460
Oy 301 AACCTGTCTCATTAACAACCAAGTAGTAGTAGTAATAGCATTAAGAAATTTGA 360
Db 40461 AACCTGTCTCATTAACAACCAAGTAGTAGTAGTAATAGCATTAAGAAATTTGA 40520
Oy 361 GTCCATTACAGATGACCATCTTAATAGATTTCTTGACCCAGTAAGCTAATGTCA 420
Db 40521 GTCCATTACAGATGACCATCTTAATAGATTTCTTGACCCAGTAAGCTAATGTCA 40580
Oy 421 TGGGAAAGGGATGGGATGTCCTAGATTAAGTAAGTGGTGAAGGCGATCTTCTCA 480
Db 40581 TGGGAAAGGGATGGGATGTCCTAGATTAAGTAAGTGGTGAAGGCGATCTTCTCA 40640
Oy 481 TTTGATTCATATGAAGAGCTGATTAAGGCCAAGAGAGTGAAGTGGACTGTGACT 540
Db 40641 TTTGATTCATATGAAGAGCTGATTAAGGCCAAGAGAGTGAAGTGGACTGTGACT 40700
Oy 541 GAAGACGTGACGCGCTTAATAACACTGGCATTTATTAACCTTAATTAACCTGGC 600
Db 40701 GAAGACGTGACGCGCTTAATAACACTGGCATTTATTAACCTTAATTAACCTGGC 40760
Oy 601 CGTTAGGTTGAAGTACCTTCAACACAGACAGAAATGCGTCTGCTCCACAG 660
Db 40761 CGTTAGGTTGAAGTACCTTCAACACAGACAGAAATGCGTCTGCTCCACAG 40820
Oy 661 GTAGGAGCACTGCTGCGAGAGAGTATTTAGTGAAGCTACCTTCAATATCTTT 720
Db 40821 GTAGGAGCACTGCTGCGAGAGAGTATTTAGTGAAGCTACCTTCAATATCTTT 40880
Oy 721 GCACATTACATACAGCTGTAAATGTGCTTAAGTCCAGTCCACAGATGGCTTTACA 780
Db 781 GCACATTACATACAGCTGTAAATGTGCTTAAGTCCAGTCCACAGATGGCTTTACA
```

```
Db 40881 GCACATTACATACAGCTGTAAATGTGCTTAAGTCCAGTCCAGATGGCTTTACA 40940
Oy 781 CCGGTTTCGCTTCCCATCTGCTGACATTTGTCAGAAACCGAAATTAAGATGGGT 840
Db 40941 CCGGTTTCGCTTCCCATCTGCTGACATTTGTCAGAAACCGAAATTAAGATGGGT 41000
Oy 841 ATTTATTTGTGTGTGAGGACACATCCAGGCTTTTCATATTCAGGACATATGTTTAC 900
Db 41001 ATTTATTTGTGTGTGAGGACACATCCAGGCTTTTCATATTCAGGACATATGTTTAC 41060
Oy 901 TTAAGTGGCTACTTCTCCAGCGTTTGAACCATTTGTTTATTTACTATTTTGTGT 960
Db 41061 TTAAGTGGCTACTTCTCCAGCGTTTGAACCATTTGTTTATTTACTATTTTGTGT 41120
Oy 961 GCATGAGTACGCAATGATATACATATGATAGAGATCATGATGCTGCTACCTCAAA 1020
Db 41121 GCATGAGTACGCAATGATATACATATGATAGAGATCATGATGCTGCTACCTCAAA 41180
Oy 1021 ATCATTTGAGATCCCGACAGATGAAGTACCGAGCGTTGTAAGTTGTTATGGGACTG 1080
Db 41181 ATCATTTGAGATCCCGACAGATGAAGTACCGAGCGTTGTAAGTTGTTATGGGACTG 41240
Oy 1081 GGAGCCAAAGCTGGGTTCTCTGCAAGAGCAGCCAGTGGCTTAACCATGGACAGCTCT 1140
Db 41241 GGAGCCAAAGCTGGGTTCTCTGCAAGAGCAGCCAGTGGCTTAACCATGGACAGCTCT 41300
Oy 1141 CTAGGCTTAAGTAACTTTTAAATATATATTTCTCAGCCGCGTGTGTGTC 1200
Db 41301 CTAGGCTTAAGTAACTTTTAAATATATATTTCTCAGCCGCGTGTGTGTC 41360
Oy 1201 ACACGCTTTAATCCAGCACTTGAAGAGCTGAGGCTGTGAATATATACACAGGCGAG 1260
Db 41361 ACACGCTTTAATCCAGCACTTGAAGAGCTGAGGCTGTGAATATATACACAGGCGAG 41420
Oy 1261 CTGGGGTCAGAGCTTGGCCCTGTTTTTTTGTGTTTTCTTAATGTGCACTGTGCTTA 1320
Db 41421 CTGGGGTCAGAGCTTGGCCCTGTTTTTTTGTGTTTTCTTAATGTGCACTGTGCTTA 41480
Oy 1321 CCTGGGTATGTCGCGTGAAGGGTGTGATCCCTGGAGTGGAGTTAAAGACAGTTG 1380
Db 41481 CCTGGGTATGTCGCGTGAAGGGTGTGATCCCTGGAGTGGAGTTAAAGACAGTTG 41540
Oy 1381 TGATCAGCTGCGCTTACAGATGTGGAATTAAGAACCCAGGTGCTCCATGAAGACAGGC 1440
Db 41541 TGATCAGCTGCGCTTACAGATGTGGAATTAAGAACCCAGGTGCTCCATGAAGACAGGC 41600
Oy 1441 AGTGTCTTAATCTTGAGCCACCCCTCCACCTGCTTTTAAGAGCTCTTAACCTTTTG 1500
Db 41601 AGTGTCTTAATCTTGAGCCACCCCTCCACCTGCTTTTAAGAGCTCTTAACCTTTTG 41660
Oy 1501 TGTAAATGGGAAGTGAAGGATCTTGCACTTAAGCAAGTGTGCTGCGCTGTACATCA 1560
Db 41661 TGTAAATGGGAAGTGAAGGATCTTGCACTTAAGCAAGTGTGCTGCGCTGTACATCA 41720
Oy 1561 CTGAGCCCGTTACCCACAGCACTAGTGAATCAAGTTAAAGGCAACACTTAACAATGACA 1620
Db 41721 CTGAGCCCGTTACCCACAGCACTAGTGAATCAAGTTAAAGGCAACACTTAACAATGACA 41780
Oy 1621 ATAGTTGATAGATTTGAATATAGTCTGAGCTATTTGTTAGCGTGAACCTTTGCTGTC 1680
Db 41781 ATAGTTGATAGATTTGAATATAGTCTGAGCTATTTGTTAGCGTGAACCTTTGCTGTC 41840
Oy 1681 TTAGCATGTGCTGAGAAAGATAGAAATGAAGACTTGAAGCTGAGTCTGGAACCCACA 1740
Db 41841 TTAGCATGTGCTGAGAAAGATAGAAATGAAGACTTGAAGCTGAGTCTGGAACCCACA 41900
Oy 1741 GAGCAGGCGAGAACCCACTCTCGAAAGTGTCTCTGAGCTTCACATACAACTTCACAT 1800
Db 41901 GAGCAGGCGAGAACCCACTCTCGAAAGTGTCTCTGAGCTTCACATACAACTTCACAT 41960
Oy 1801 AATAGTTACATGATATATATATATATATTTCTTTTAAAGGTATATGTTGGAGGGA 1860
Db 41961 AATAGTTACATGATATATATATATATATTTCTTTTAAAGGTATATGTTGGAGGGA 42020
```

OY 1861 GAGATGGCTACGCTCCAGAGACCTTGGCTTTCGACAGAGACCTAGATTTCAGTTCCC 1920
|||||
DB 42021 GAGATGGCTACGCTCCAGAGACCTTGGCTTTCGACAGAGACCTAGATTTCAGTTCCC 42080
OY 1921 AGCACTCATATGGTGGCTCACAAGCCATCTGTAAATCCAGTTCCAGAGGGTCCACACCT 1980
|||||
DB 42081 AGCACTCATATGGTGGCTCACAAGCCATCTGTAAATCCAGTTCCAGAGGGTCCACACCT 42140
OY 1981 CTTTGGGCTCCACAGACCCATATCATATGATACAGACATATGACAGGCAAAACACC 2040
|||||
DB 42141 CTTTGGGCTCCACAGACCCATATCATATGATACAGACATATGACAGGCAAAACACC 42200
OY 2041 CATACACATATAATATAAAGAACTTAAAGTGATCTGTGTAAACCTTGTCT 2100
|||||
DB 42201 CATACACATATAATATAAAGAACTTAAAGTGATCTGTGTAAACCTTGTCT 42260
OY 2101 TACACATGCTGATTTGAAGACATGTACACGCACACTGAAGAGGATCTGGGGCTGGAG 2160
|||||
DB 42261 TACACATGCTGATTTGAAGACATGTACACGCACACTGAAGAGGATCTGGGGCTGGAG 42320
OY 2161 AGATGGCTACGGGGTTAAAGACATGCTGCTTTCGGAAGGAGAGTCCGTGAATTCAAT 2220
|||||
DB 42321 AGATGGCTACGGGGTTAAAGACATGCTGCTTTCGGAAGGAGAGTCCGTGAATTCAAT 42380
OY 2221 CCTAGCAACCATGCTGGCTCACAACCATCATATGATGATCTGACACCTTCTGGT 2280
|||||
DB 42381 CCTAGCAACCATGCTGGCTCACAACCATCATATGATGATCTGACACCTTCTGGT 42440
OY 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTTATGATATATAATAATCTTTT 2340
|||||
DB 42441 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTTATGATATATAATAATCTTTT 42500
OY 2341 TTTAAAAAATGAAGGAGATCTGAGACCTCAAAAGATATATGAGAGTACTCAG 2400
|||||
DB 42501 TTTAAAAAATGAAGGAGATCTGAGACCTCAAAAGATATATGAGAGTACTCAG 42560
OY 2401 GGTGATTATCTATCTCTGAGATTCTTTCGCTTGGCTTGCATGCTGGTGAGACAG 2460
|||||
DB 42561 GGTGATTATCTATCTCTGAGATTCTTTCGCTTGGCTTGCATGCTGGTGAGACAG 42620
OY 2461 CCCCTTTTATCTACAGAAGGGGCTCTACATTAATTTGAAACAAAAGCACTCTCAGT 2520
|||||
DB 42621 CCCCTTTTATCTACAGAAGGGGCTCTACATTAATTTGAAACAAAAGCACTCTCAGT 42680
OY 2521 ATGTTTACTGCTGCTGCTACTATGAGCAAGGGGACGGGGGCGG--CACACACACAC 2578
|||||
DB 42681 ATGTTTACTGCTGCTGCTACTATGAGCAAGGGGACGGGGGCGG--CACACACACAC 42740
OY 2579 ACACACACACACACACACACACACACATCTCCAGAGCTCTTGGAGAGT 2638
|||||
DB 42741 ACACACACACACACACACACACACACATCTCCAGAGCTCTTGGAGAGT 42800
OY 2639 CAAGAAGAGGCTCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGACACGAT 2698
|||||
DB 42801 CAAGAAGAGGCTCCCTCAAAACAGATCTTTCATCTTCCCTCTTAAAGAGACACGAT 42860
OY 2699 CCAAGGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGGGGAAGGAGGATGTTT 2758
|||||
DB 42861 CCAAGGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGGGGAAGGAGGATGTTT 42920
OY 2759 CCAGAGACCTTACAGCAGAGGAGCAGAGGAGATCCCAAGTCTCCAGGGCAGGAGGTGA 2818
|||||
DB 42921 CCAGAGACCTTACAGCAGAGGAGCAGAGGAGATCCCAAGTCTCCAGGGCAGGAGGTGA 42980
OY 2819 GGGCTTGTTCGAGAGAGAGGAGGAGCAGACAGGCTTCAAGAGCACAGGTTTATGG 2878
|||||
DB 42981 GGGCTTGTTCGAGAGAGAGGAGGAGCAGACAGGCTTCAAGAGCACAGGTTTATGG 43040
OY 2879 CAGCTATATAAAGTGAGGCTGCTGCTCACTCAGAAAGAGAGAAAGGAAAGGCGCTT 2938
|||||
DB 43041 CAGCTATATAAAGTGAGGCTGCTGCTCACTCAGAAAGAGAGAAAGGAAAGGCGCTT 43100

OY 2939 GTGCCACTGAGCAGAGGCTCATGCTAGTAGAGAGATCTGCAGGGGTCAGAGACCC 2998
|||||
DB 43101 GTGCCACTGAGCAGAGGCTCATGCTAGTAGAGAGATCTGCAGGGGTCAGAGACCC 43160
OY 2999 ACCTGTCTGTCCCAAGGAGAACCCCAAGTGTGAATCTTGCCCTTGGGTGCTGAGTTCCAGC 3058
|||||
DB 43161 ACCTGTCTGTCCCAAGGAGAACCCCAAGTGTGAATCTTGCCCTTGGGTGCTGAGTTCCAGC 43220
OY 3059 TACAGAGCCCGAGAGTCTCTACTCATCTCCCAACATGCTCCCGCCCGCCAGCAC 3118
|||||
DB 43221 TACAGAGCCCGAGAGTCTCTACTCATCTCCCAACATGCTCCCGCCCGCCAGCAC 43280
OY 3119 CACCCCGCATCCCGGCTGCTTCTAGGGCTGAGAGGTGGCCAGCCCTGTGGGGTT 3178
|||||
DB 43281 CACCCCGCATCCCGGCTGCTTCTAGGGCTGAGAGGTGGCCAGCCCTGTGGGGTT 43340
OY 3179 GCTTACCTGACAGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCCTGAAGCTGCA 3238
|||||
DB 43341 GCTTACCTGACAGTAGAGCCCAAGTCTTAGCCGGAAGTGCACCCCATCCCTGAAGCTGCA 43400
OY 3239 GAGCCAGGGGGGGGACACGGCAGCTCAGGCTGTCAGGCTGTGCTGGGCTTACGTTTC 3298
|||||
DB 43401 GAGCCAGGGGGGGGACACGGCAGCTCAGGCTGTCAGGCTGTGCTGGGCTTACGTTTC 43460
OY 3299 CCAGGGACCTGGGGACCTACTTCCCCACCCCCCATCTCTCTGCGGGCCATCTT 3358
|||||
DB 43461 CCAGGGACCTGGGGACCTACTTCCCCACCCCCCATCTCTCTGCGGGCCATCTT 43520
OY 3359 CCCTTATATGTTGAAGAAATCTTGGGGGGGGGGGGTGTGTGAGCAAAAGTCTGT 3418
|||||
DB 43521 CCCTTATATGTTGAAGAAATCTTGGGGGGGGGGGGTGTGTGAGCAAAAGTCTGT 43578
OY 3419 GGGTCTCCGTCAGCCAGCTTGGCCACAACCTTCTAAGATCTCCAGAGTGGTGGCTGCT 3478
|||||
DB 43579 GGGTCTCCGTCAGCCAGCTTGGCCACAACCTTCTAAGATCTCCAGAGTGGTGGCTGCT 43638
OY 3479 TCCAGACAGTAAAGCAATTTGGGTGGGACACATGTTGACCAAGAGTGGTTGAGAGGAC 3538
|||||
DB 43639 TCCAGACAGTAAAGCAATTTGGGTGGGACACATGTTGACCAAGAGTGGTTGAGAGGAC 43698
OY 3539 AGGTCCTTGTCTCTCTGACAGCTGTCTTCTGTATGACACTTGGTATTAAGTTGGG 3598
|||||
DB 43699 AGGTCCTTGTCTCTCTGACAGCTGTCTTCTGTATGACACTTGGTATTAAGTTGGG 43758
OY 3599 GGTGAGGTAAAGGCTCTGAACCTTGAAGAAGCAAGAGGACAGAGCTCTCTTGGG 3658
|||||
DB 43759 GGTGAGGTAAAGGCTCTGAACCTTGAAGAAGCAAGAGGACAGAGCTCTCTTGGG 43818
OY 3659 CTTCAATGAAGGAAGTTACAGACCCCTTCTCTGAAGTCACTTGGCTTCACTGTGT 3718
|||||
DB 43819 CTTCAATGAAGGAAGTTACAGACCCCTTCTCTGAAGTCACTTGGCTTCACTGTGT 43878
OY 3719 ACATTTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTTCAATTAATTCAGAGACA 3778
|||||
DB 43879 ACATTTCCCTGGGACCAAGTGGCTCTGGGACTCAGATTTTCAATTAATTCAGAGACA 43938
OY 3779 GTCTGAGACTTGGACTCCGCTGCTATTTACTACTTCTCTGCTGCTCTCATTTCTGT 3838
|||||
DB 43939 GTCTGAGACTTGGACTCCGCTGCTATTTACTACTTCTCTGCTGCTCTCATTTCTGT 43998
OY 3839 GTTCATGCTTACACATCTGAATGTTTCTTGTGTCCATCTTCCCTGACATCTCTGG 3898
|||||
DB 43999 GTTCATGCTTACACATCTGAATGTTTCTTGTGTCCATCTTCCCTGACATCTCTGG 44058
OY 3899 GAGTGTCTTCTTGGCAGATGATCTTGGGATGTAGCTGACCCAGCAGAGAGAGG 3958
|||||
DB 44059 GAGTGTCTTCTTGGCAGATGATCTTGGGATGTAGCTGACCCAGCAGAGAGAGG 44118
OY 3959 GGAGGTGAGAGGCTGTCTAGGCGCTTATTAAGGCTGGAATACCCCTTTCATAGAA 4018
|||||
DB 44119 GGAGGTGAGAGGCTGTCTAGGCGCTTATTAAGGCTGGAATACCCCTTTCATAGAA 44178
OY 4019 ATGGCCCTCTCATTTTTCGGTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGC 4078

* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 46829: contig of 46829 bp in length
 46830 46929: gap of 100 bp
 46930 49621: contig of 2692 bp in length
 49622 49721: gap of 100 bp
 49722 70009: contig of 20288 bp in length
 70010 70109: gap of 100 bp
 70110 77196: contig of 7087 bp in length
 77197 77296: gap of 100 bp
 77297 99453: contig of 22157 bp in length
 99454 99553: gap of 100 bp
 99554 102879: contig of 3326 bp in length
 102880 102979: gap of 100 bp
 102980 120352: contig of 17373 bp in length
 120353 120452: gap of 100 bp
 120453 131546: contig of 11094 bp in length
 131547 131646: gap of 100 bp
 131647 149109: contig of 17463 bp in length.

Location/Qualifiers
 1. 149109
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone_11b="RP32-328P7"
 /clone_11b="RP32-328P7"

BASE COUNT 39036 a 34316 c 35185 g 39719 t 853 others

ORIGIN

Query Match 1.1% Score 96; DB 2; Length 149109;
 Best Local Similarity 100.0%; Pred. No. 9e-41;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5953 GAGTTTACCAAGAGGTAAAGCCAGCATTTATTCACAAAGACATGCTCAAAATCAAG 6012
 DB 102361 GAGTTTACCAAGAGGTAAAGCCAGCATTTATTCACAAAGACATGCTCAAAATCAAG 102420

OY 6013 TGCAGAGAGATTAAGTCTGAGAGATGGGCTGTCA 6048
 DB 102421 TGCAGAGAGATTAAGTCTGAGAGATGGGCTGTCA 102456

RESULT 4
 AC115173 159492 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-286017, *** SEQUENCING IN PROGRESS
 DEFINITION *** 46 unordered pieces.
 AC115173
 VERSION AC115173.2 GI:21738030
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 159492)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alstrooms,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Blinze,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunyah,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H., Douthwaite,K.J., Drepper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabrieli,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Hayak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Medora,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogum,W., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Plickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Wetstock,G. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 159492)
 Worley,K.C.

Direct Submission
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 159492)
 Worley,K.C.

Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:19482237.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GODN
 Center clone name: CH230-286017
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 124398 bases at least Q40
 Consensus quality: 128059 bases at least Q30
 Consensus quality: 131015 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1014: contig of 1014 bp in length
 * 1015 1114: gap of unknown length
 * 1115 2395: contig of 1281 bp in length
 * 2396 2495: gap of unknown length

| | | |
|-------|--------|-----------------------------|
| 2496 | 3600: | contig of 1105 bp in length |
| 3601 | 3700: | gap of unknown length |
| 3701 | 5021: | contig of 1321 bp in length |
| 5022 | 5121: | gap of unknown length |
| 5122 | 6727: | contig of 1606 bp in length |
| 6728 | 6827: | gap of unknown length |
| 6828 | 8608: | contig of 1781 bp in length |
| 8609 | 8708: | gap of unknown length |
| 8709 | 10003: | contig of 1295 bp in length |
| 10004 | 10103: | gap of unknown length |
| 10104 | 11211: | contig of 1108 bp in length |
| 11212 | 11311: | gap of unknown length |
| 11312 | 12518: | contig of 1207 bp in length |
| 12519 | 12618: | gap of unknown length |
| 12619 | 13843: | contig of 1225 bp in length |
| 13844 | 13943: | gap of unknown length |
| 13944 | 15404: | contig of 1461 bp in length |
| 15405 | 15504: | gap of unknown length |
| 15505 | 17647: | contig of 2143 bp in length |
| 17648 | 17747: | gap of unknown length |
| 17748 | 19706: | contig of 1959 bp in length |
| 19707 | 19806: | gap of unknown length |
| 19807 | 21576: | contig of 1770 bp in length |
| 21577 | 21677: | gap of unknown length |
| 21677 | 22906: | contig of 1230 bp in length |
| 22907 | 23006: | gap of unknown length |
| 23007 | 25086: | contig of 2080 bp in length |
| 25087 | 25186: | gap of unknown length |
| 25187 | 26832: | contig of 1646 bp in length |
| 26833 | 26932: | gap of unknown length |
| 26932 | 28710: | contig of 1778 bp in length |
| 28711 | 28810: | gap of unknown length |
| 28811 | 30606: | contig of 1796 bp in length |
| 30607 | 30706: | gap of unknown length |
| 30707 | 33113: | contig of 2407 bp in length |
| 33114 | 33213: | gap of unknown length |
| 33214 | 35361: | contig of 2148 bp in length |
| 35362 | 35461: | gap of unknown length |
| 35462 | 37520: | contig of 2059 bp in length |
| 37521 | 37620: | gap of unknown length |
| 37621 | 40471: | contig of 2851 bp in length |
| 40472 | 40571: | gap of unknown length |
| 40572 | 42584: | contig of 2013 bp in length |
| 42585 | 42684: | gap of unknown length |
| 42685 | 45547: | contig of 2863 bp in length |
| 45548 | 45647: | gap of unknown length |
| 45648 | 47838: | contig of 2191 bp in length |
| 47839 | 47938: | gap of unknown length |
| 47939 | 51229: | contig of 3291 bp in length |
| 51230 | 51329: | gap of unknown length |
| 51330 | 54005: | contig of 2676 bp in length |
| 54006 | 54105: | gap of unknown length |
| 54106 | 55683: | contig of 2578 bp in length |
| 55684 | 55783: | gap of unknown length |
| 55784 | 55961: | contig of 2578 bp in length |
| 55962 | 59461: | gap of unknown length |
| 59462 | 61588: | contig of 2127 bp in length |
| 61589 | 61888: | gap of unknown length |
| 61890 | 65712: | contig of 4024 bp in length |
| 65713 | 65812: | gap of unknown length |
| 65813 | 65939: | contig of 3527 bp in length |
| 65940 | 66340: | gap of unknown length |
| 66341 | 72570: | contig of 3131 bp in length |
| 72571 | 72670: | gap of unknown length |
| 72671 | 76898: | contig of 4228 bp in length |
| 76899 | 76998: | gap of unknown length |
| 76999 | 80760: | contig of 3762 bp in length |
| 80761 | 80860: | gap of unknown length |
| 80861 | 86313: | contig of 5453 bp in length |
| 86314 | 86413: | gap of unknown length |
| 86414 | 92231: | contig of 5818 bp in length |
| 92232 | 92331: | gap of unknown length |
| 92332 | 92331: | gap of unknown length |
| 92333 | 96961: | contig of 7361 bp in length |

| | | | |
|-------------|--------------------------------|---|---|
| FEATURES | source | 1. 159492 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-286017" | 39083 a 36447 c 36055 g 40464 t 7443 others |
| BASE COUNT | ORIGIN | | |
| Query Match | Best local Similarity | 1.1%: Score 96: DB 2: Length 159492: | |
| Matches | 96: Conservative | 100.0%: Pred.No. 9e-41: Mismatches 0: Indels 0: Gaps 0: | |
| Oy | 5953 | GAGTTTACAGAGGTAAGGCCAGCCATTTATTATCCAGACATGACATCAAAATCAAG | 6012 |
| Db | 27228 | GAGTTTACAGAGGTAAGGCCAGCCATTTATTATCCAGACATGACATCAAAATCAAG | 27169 |
| Oy | 6013 | TGCAAGAGAGATTAGCTGAGAGATGGGGCTCA | 6048 |
| Db | 27168 | TGCAAGAGAGATTAGCTGAGAGATGGGGCTCA | 27133 |
| RESULT 5 | AC127697 | | |
| LOCUS | AC127697 | | |
| DEFINITION | | Mus musculus chromosome UNK clone RP24-83110, WORKING DRAFT | |
| ACCESSION | AC127697 | | |
| VERSION | AC127697.2 | | |
| KEYWORDS | HTG; HTGS; PHASE1; HTGS_DRAFT. | | |
| SOURCE | | Mus musculus. | |
| ORGANISM | | Mus musculus. | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| AUTHORS | | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| TITLE | | 1 (bases 1 to 222185) | |
| REFERENCE | | McPherson,J.D. and Waterston,R.H. | |
| JOURNAL | | The sequence of Mus musculus clone | |
| AUTHORS | | Unpublished | |
| TITLE | | 2 (bases 1 to 222185) | |
| JOURNAL | | McPherson,J.D. and Waterston,R.H. | |
| REFERENCE | | Direct Submission | |
| AUTHORS | | Submitted (18-JUL-2002) Genome Sequencing Center, 4444 Forest Park | |
| TITLE | | Parkway, St. Louis, MO 63108, USA | |
| JOURNAL | | 3 (bases 1 to 222185) | |
| REFERENCE | | McPherson,J.D. and Waterston,R.H. | |
| AUTHORS | | Direct Submission | |
| TITLE | | Submitted (22-JUL-2002) Genome Sequencing Center, 4444 Forest Park | |
| JOURNAL | | Parkway, St. Louis, MO 63108, USA | |
| COMMENT | | On Jul 21, 2002 this sequence version replaced gi:21903681. | |

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Contact: submissions@wustl.edu

----- Project Information -----

Center project name: M_BB0083110

----- Summary Statistics -----

Sequencing vector: M13; 0%

```
Sequencing vector: plasmid; 100%
Chemistry: Dye-Primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 217367 bases at least Q40
Consensus quality: 218557 bases at least Q30
Consensus quality: 219259 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1042: contig of 1042 bp in length
* 1043 1142: gap of unknown length
* 1143 3141: contig of 1999 bp in length
* 3142 3241: gap of unknown length
* 3242 7211: contig of 3970 bp in length
* 7212 7311: gap of unknown length
* 7312 10408: contig of 3097 bp in length
* 10409 10508: gap of unknown length
* 10509 17015: contig of 6507 bp in length
* 17016 17115: gap of unknown length
* 17116 27558: contig of 10443 bp in length
* 27559 27658: gap of unknown length
* 27659 41472: contig of 13814 bp in length
* 41473 41572: gap of unknown length
* 41573 58617: contig of 17045 bp in length
* 58618 83266: contig of 24549 bp in length
* 83267 83366: gap of unknown length
* 83367 130000: contig of 46634 bp in length
* 130001 130100: gap of unknown length
* 130101 172736: contig of 42636 bp in length
* 172737 172836: gap of unknown length
* 172837 221107: contig of 48271 bp in length
* 221108 221208: gap of unknown length
* 221209 222185: contig of 978 bp in length.
Location/Qualifiers
1. 222185
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-83110"
misc_feature
1. 1042
/note="assembly_name:Contig25"
misc_feature
1143. 3141
/note="assembly_name:Contig26"
misc_feature
3242. 7211
/note="assembly_name:Contig27"
misc_feature
7312. 10408
/note="assembly_name:Contig28"
misc_feature
10509. 17015
/note="assembly_name:Contig29"
misc_feature
17116. 27558
/note="assembly_name:Contig30"
misc_feature
27659. 41472
/note="assembly_name:Contig31"
misc_feature
41573. 58617
/note="assembly_name:Contig32"
misc_feature
58718. 83266
/note="assembly_name:Contig33"
misc_feature
83367. 130000
/note="assembly_name:Contig34"
misc_feature
130101. 172736
/note="assembly_name:Contig35"
misc_feature
172837. 222107
/note="assembly_name:Contig36"
misc_feature
221208. 222185
/note="assembly_name:Contig37"
-----
BASE COUNT 67728 a 41550 c 41568 g 70139 t 1200 others
ORIGIN
Query Match 0.9%; Score 82; DB 2; Length 222185;
Best Local Similarity 100.0%; Pred. No. 4.3e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4593 GCAGTGTGTCACAGCCCTTTATCCACACTTGGAGCAGAGCAGATTTCG 4652
DB 26396 GCAGTGTGTCACAGCCCTTTATCCACACTTGGAGCAGAGCAGATTTCG 26455
Qy 4653 AGTTCAGGCCAGCCCTGTCTA 4674
DB 26456 AGTTCAGGCCAGCCCTGTCTA 26477
-----
RESULT 6
AC079490/c 226909 bp DNA linear HTG 02-SEP-2000
LOCUS Mus musculus clone RP23-190A2, WORKING DRAFT SEQUENCE, 53 unordered
DEFINITION pieces.
AC079490
VERSION AC079490.1 GI:9964855
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 226909)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 226909)
REFERENCE DOE Joint Genome Institute.
TITLE Direct Submission.
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1809834
Center clone name: RPC1-23_190A2
-----
Summary Statistics
Consensus quality: 180161 bases at least Q40
Consensus quality: 198601 bases at least Q30
Consensus quality: 203733 bases at least Q20
Estimated insert size: 209000; agarose-fp estimation
Estimated insert size: 221709; sum-of-contigs estimation
Quality coverage: 6.2 in Q20 bases; agarose-fp estimation
Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1120: contig of 1120 bp in length
* 1121 1220: gap of unknown length
* 1221 1221: contig of 1354 bp in length
* 1222 2574: gap of unknown length
* 2575 2675: gap of unknown length
* 2676 4062: contig of 1388 bp in length
* 4063 4163: gap of unknown length
* 4164 5667: contig of 1504 bp in length
* 5668 5767: gap of unknown length
* 5768 6895: contig of 1129 bp in length
* 6896 6996: gap of unknown length
* 6997 8025: contig of 1030 bp in length
```


* 8026 8125: gap of unknown length
* 9388: contig of 1263 bp in length
* 9389 9488: gap of unknown length
* 9489 10710: contig of 1222 bp in length
* 10711 10810: gap of unknown length
* 10811 11811: contig of 1001 bp in length
* 11812 11911: gap of unknown length
* 11912 12965: contig of 1054 bp in length
* 12966 13065: gap of unknown length
* 13066 14268: contig of 1203 bp in length
* 14269 14368: gap of unknown length
* 14369 15846: contig of 1478 bp in length
* 15847 17219: contig of 1273 bp in length
* 15947 17319: gap of unknown length
* 17220 18672: contig of 1353 bp in length
* 18673 18772: gap of unknown length
* 18773 19999: contig of 1227 bp in length
* 20000 20099: gap of unknown length
* 20100 21459: contig of 1360 bp in length
* 21460 21559: gap of unknown length
* 21560 22622: contig of 1063 bp in length
* 22623 22722: gap of unknown length
* 22723 23726: contig of 1004 bp in length
* 23727 23826: gap of unknown length
* 23827 25149: contig of 1323 bp in length
* 25150 25249: gap of unknown length
* 25250 26839: contig of 1590 bp in length
* 26840 26939: gap of unknown length
* 26940 28630: contig of 1691 bp in length
* 28631 28730: gap of unknown length
* 28731 30265: contig of 1535 bp in length
* 30266 30365: gap of unknown length
* 30366 31703: contig of 1338 bp in length
* 31704 31803: gap of unknown length
* 31804 32951: contig of 1148 bp in length
* 32952 33051: gap of unknown length
* 33052 35329: contig of 2278 bp in length
* 35330 35429: gap of unknown length
* 35430 38496: contig of 3067 bp in length
* 38497 38596: gap of unknown length
* 38597 40521: contig of 1925 bp in length
* 40522 40621: gap of unknown length
* 40622 42212: contig of 1591 bp in length
* 42213 42312: gap of unknown length
* 42313 44441: contig of 2129 bp in length
* 44442 44541: gap of unknown length
* 44542 46397: contig of 1856 bp in length
* 46398 46497: gap of unknown length
* 46498 48685: contig of 2188 bp in length
* 48686 48785: gap of unknown length
* 48786 50811: contig of 2026 bp in length
* 50812 50911: gap of unknown length
* 50912 52640: contig of 1729 bp in length
* 52641 52740: gap of unknown length
* 52741 53952: contig of 1212 bp in length
* 53953 54052: gap of unknown length
* 54053 58018: contig of 3966 bp in length
* 58019 58118: gap of unknown length
* 58119 61436: contig of 3318 bp in length
* 61437 61537: gap of unknown length
* 61538 64032: contig of 2496 bp in length
* 64033 64132: gap of unknown length
* 64133 66752: contig of 2620 bp in length
* 66753 66852: gap of unknown length
* 66853 70097: contig of 3245 bp in length
* 70098 70197: gap of unknown length
* 70198 73503: contig of 3306 bp in length
* 73504 73603: gap of unknown length
* 73604 80346: contig of 6743 bp in length
* 80347 80446: gap of unknown length
* 80447 86466: contig of 6020 bp in length
* 86467 86566: gap of unknown length

* 86567 92278: contig of 5712 bp in length
* 92279 92378: gap of unknown length
* 92379 98874: contig of 6496 bp in length
* 98875 98974: gap of unknown length
* 98975 106050: contig of 7076 bp in length
* 106051 106150: gap of unknown length
* 106151 113625: contig of 7475 bp in length
* 113626 113725: gap of unknown length
* 113726 129075: contig of 15350 bp in length
* 129076 129175: gap of unknown length
* 129176 141800: contig of 12625 bp in length
* 141801 141900: gap of unknown length
* 141901 155581: contig of 13681 bp in length
* 155582 170324: gap of unknown length
* 155682 170324: contig of 14643 bp in length
* 170325 170424: gap of unknown length
* 170425 184681: contig of 14257 bp in length
* 184682 206159: contig of 21378 bp in length
* 206160 206259: gap of unknown length
* 206260 226909: contig of 20650 bp in length.
FEATURES
source
1. .265909
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-190A2"
/clone_lib="RPCT mouse BAC library 23"
BASE COUNT 52802 a 57780 c 58798 g 52284 t 5245 others
ORIGIN
Query Match 0.9%: Score 82: DB 2: Length 226909;
Best Local Similarity 100.0%: Pred. No. 4.3e-33;
Matches 82: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 4593 GCAGTGTGGCAGACGCCCTTATTCACGACCTTGAGGAGGAGGAGGAGGAGATTCTG 4652
|||||
Db 35815 GCAGGTGGGACACGCCCTTATTCACGACCTTGAGGAGGAGGAGGAGGAGATTCTG 35756
QY 4653 AGTTCAGGCCACGCTGTGCTTA 4674
|||||
Db 35755 AGTTCAGGCCACGCTGTGCTTA 35734
RESULT 7
AC093351/c 239077 bp DNA linear HTG 07-NOV-2001
LOCUS
DEFINITION
AC093351
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
TITLE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
AC093351
Mus musculus clone RP23-16L19, WORKING DRAFT SEQUENCE, 12 unordered
pieces.
AC093351
AC093351.1 GI:15213896
HTG: HTGS_PHASE1: HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 239077)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-16L19
Unpublished
2 (bases 1 to 239077)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, T., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deavrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lamazares, R., Landers, T., Lehoczkay, J., Levine, R., Liu, G.,
MacLean, C., MacDonald, P., Major, J., Margus, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

TITLE
JOURNAL
COMMENT

Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhahang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schnupack, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (20-AUG-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L13561

Center clone name: 16_L_19

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 234078 bases at least Q40

Consensus quality: 237194 bases at least Q20

Insert size: 222000; agarose-fp

Insert size: 237977; sum-of-contigs

Quality coverage: 11.5 in Q20 bases; agarose-fp
Quality coverage: 10.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 12 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 5895: contig of 5895 bp in length
5896 5995: gap of 100 bp
5996 6105: contig of 110 bp in length
6106 6205: gap of 100 bp
6206 6873: contig of 668 bp in length
6874 6973: gap of 100 bp
6974 7992: contig of 1019 bp in length
7993 8092: gap of 100 bp
8093 10126: contig of 2034 bp in length
10127 10226: gap of 100 bp
10227 16035: contig of 5809 bp in length
16036 16135: gap of 100 bp
16136 30334: contig of 14199 bp in length
30335 30434: gap of 100 bp
30435 59946: contig of 29512 bp in length
59947 60046: gap of 100 bp
60047 79068: contig of 19022 bp in length
79069 79168: gap of 100 bp
79169 122677: contig of 43509 bp in length
122678 122777: gap of 100 bp
122778 176372: contig of 53595 bp in length
176373 176473: gap of 100 bp
176473 239077: contig of 62605 bp in length.

FEATURES

source

1. 239077

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="RP23-16L19"

/clone_lib="RPCT-23 Female Mouse BAC"

misc_feature

1. 5895

/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
5996..6105
/note="assembly_fragment"
misc_feature
6206..6873
/note="assembly_fragment"
misc_feature
6974..7992
/note="assembly_fragment"
misc_feature
8093..10126
/note="assembly_fragment"
misc_feature
10227..16035
/note="assembly_fragment"
misc_feature
16136..30334
/note="assembly_fragment"
misc_feature
30435..59946
/note="assembly_fragment"
misc_feature
60047..79068
/note="assembly_fragment"
misc_feature
79169..122677
/note="assembly_fragment"
misc_feature
122778..176372
/note="assembly_fragment"
misc_feature
176473..239077
/note="assembly_fragment"

BASE COUNT 65130 a 55831 c 54944 g 62063 t 1109 others
ORIGIN

Query Match 0.9%; Score 82; DB 2; Length 239077;
Best Local Similarity 100.0%; Pred. No. 4,36-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4638 GCAGCAGATTTCGATTCAGAGCCAGCCGCTGCTATAGAGTCCAGACAGCA 4697
Db 42272 GCAGCAGATTTCGATTCAGAGCCAGCCGCTGCTATAGAGTCCAGACAGCA 42213
OY 4698 GGGCTACACAGAGAAACCTGT 4719
Db 42212 GGGCTACACAGAGAAACCTGT 42191

RESULT 8

AC073784 235302 bp DNA linear HTG 29-JUN-2000
LOCUS Mus musculus clone RP23-412G23, WORKING DRAFT SEQUENCE, 46
DEFINITION Unordered pieces.
ACCESSION AC073784.1 GI:8810401
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 235302)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing Of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 235302)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1895247
Center clone name: RPCT-23_412G23

Summary Statistics
Consensus quality: 197724 bases at least Q40

Consensus quality: 214621 bases at least Q30
 Consensus quality: 218454 bases at least Q20
 Estimated insert size: 203000; agarose-fp estimation
 Estimated insert size: 230802; sum-of-contris estimation
 Quality coverage: 7.31 in Q20 bases; agarose-fp estimation
 Quality coverage: 6.43 in Q20 bases; sum-of-contris estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 46 contris. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contris are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1442: contig of 1442 bp in length
 * 1443 1542: gap of unknown length
 * 1543 2717: contig of 1175 bp in length
 * 2718 2817: gap of unknown length
 * 2818 3834: contig of 1017 bp in length
 * 3835 3934: gap of unknown length
 * 3935 4966: contig of 1032 bp in length
 * 4967 5066: gap of unknown length
 * 5067 6129: contig of 1063 bp in length
 * 6130 6229: gap of unknown length
 * 6230 7407: contig of 1178 bp in length
 * 7408 7507: gap of unknown length
 * 7508 8550: contig of 1043 bp in length
 * 8551 8650: gap of unknown length
 * 8651 9654: contig of 1004 bp in length
 * 9655 9754: gap of unknown length
 * 9755 11221: contig of 1467 bp in length
 * 11222 11321: gap of unknown length
 * 11322 12433: contig of 1112 bp in length
 * 12434 12533: gap of unknown length
 * 12534 13942: contig of 1409 bp in length
 * 13943 14042: gap of unknown length
 * 14043 15118: contig of 1076 bp in length
 * 15119 15218: gap of unknown length
 * 15219 17182: contig of 1964 bp in length
 * 17183 17282: gap of unknown length
 * 17283 18452: contig of 1170 bp in length
 * 18453 18552: gap of unknown length
 * 18553 20400: contig of 1848 bp in length
 * 20401 20500: gap of unknown length
 * 20501 21738: contig of 1238 bp in length
 * 21739 21838: gap of unknown length
 * 21839 23128: contig of 1290 bp in length
 * 23129 23228: gap of unknown length
 * 23229 24987: contig of 1759 bp in length
 * 24988 25087: gap of unknown length
 * 25088 26474: contig of 1387 bp in length
 * 26475 26574: gap of unknown length
 * 26575 28321: contig of 1747 bp in length
 * 28322 28421: gap of unknown length
 * 28422 30965: contig of 2544 bp in length
 * 30966 31065: gap of unknown length
 * 31066 34079: contig of 3014 bp in length
 * 34080 34179: gap of unknown length
 * 34180 36552: contig of 2373 bp in length
 * 36553 36652: gap of unknown length
 * 36653 38883: contig of 2231 bp in length
 * 38884 41860: gap of unknown length
 * 41861 41960: gap of unknown length
 * 41961 45013: contig of 3053 bp in length
 * 45014 45113: gap of unknown length
 * 45114 47059: contig of 1946 bp in length
 * 47060 47159: gap of unknown length
 * 47160 50539: contig of 3380 bp in length
 * 50540 52880: gap of unknown length
 * 52881 52980: contig of 2241 bp in length
 * 52981 55866: gap of unknown length
 * 55866: contig of 2886 bp in length

55867 55966: gap of unknown length
 * 55967 59063: contig of 3097 bp in length
 * 59064 59163: gap of unknown length
 * 59164 63021: contig of 3888 bp in length
 * 63022 63121: gap of unknown length
 * 63122 69865: contig of 6744 bp in length
 * 69866 69965: gap of unknown length
 * 69966 78312: contig of 8347 bp in length
 * 78313 78412: gap of unknown length
 * 78413 85315: contig of 6903 bp in length
 * 85316 85415: gap of unknown length
 * 85416 93535: contig of 8120 bp in length
 * 93536 93635: gap of unknown length
 * 93636 104777: contig of 11142 bp in length
 * 104778 104877: gap of unknown length
 * 104878 116166: contig of 11289 bp in length
 * 116167 116266: gap of unknown length
 * 116267 128380: contig of 12114 bp in length
 * 128381 128480: gap of unknown length
 * 128481 138205: contig of 9725 bp in length
 * 138206 138305: gap of unknown length
 * 138306 152565: contig of 14260 bp in length
 * 152566 152665: gap of unknown length
 * 152666 165170: contig of 12505 bp in length
 * 165171 165270: gap of unknown length
 * 165271 177903: contig of 12633 bp in length
 * 177904 178003: gap of unknown length
 * 178004 190290: contig of 12287 bp in length
 * 190291 190390: gap of unknown length
 * 190391 211048: contig of 20658 bp in length
 * 211049 211148: gap of unknown length
 * 211149 235302: contig of 24154 bp in length.

FEATURES
 source 1..235302
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RP23-412G23"
 /clone_lib="RPCI mouse BAC library 23"
 BASE COUNT 57053 a 57886 c 58312 g 57537 t 4514 others
 ORIGIN

Query Match 0.9%; Score 80; DB 2; Length 235302;
 Best Local Similarity 100.0%; Pred. No. 5.4e-32;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2202 GAAGGTCCTGAGTCAATCCCTAGCAACCATGCTGCTACACCATCCATATGAGA 2261
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 119422 GAAGGTCCTGAGTCAATCCCTAGCAACCATGCTGCTACACCATCCATATGAGA 119481
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

OY 2262 TCTGACACCCCTCTTGCTG 2281
 ||||||||||||||||||||||||||||
 DB 119482 TCTGACACCCCTCTTGCTG 119501

RESULT 9
 AL589737/c
 LOCUS 202855 bp DNA linear ROD 29-JUN-2002
 DEFINITION Mouse DNA sequence from clone Rp23-45L14 on chromosome 13, complete
 sequence.
 ACCESSION AL589737
 VERSION AL589737.11 GI:17976581
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 Bates, K.
 REFERENCE Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 COMMENT On Dec 21, 2001 this sequence version replaced gi:17939723.

[illegible][illegible]

Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M.,
McKean,P., McKernan,K., Meldrum,J., Meneus,L., Mithova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Noddu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strause,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
Submitted (09-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 174982)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C.,
McCarthy,M., McKean,P., McKernan,K., Meldrum,J., Meneus,L.,
Mithova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Noddu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strause,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (03-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 3, 2002 this sequence version replaced 91:18642805.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L22522

Center clone name: 237_L-12

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 167618 bases at least Q40

Consensus quality: 171113 bases at least Q30

Consensus quality: 172140 bases at least Q20

Insert size: 178000; agarose-fp

Insert size: 172582; sum-of-contigs

Quality coverage: 5.3 in Q20 bases; agarose-fp

Quality coverage: 5.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 1803: contig of 1803 bp in length
1804 1903: gap of 100 bp
1904 2784: contig of 881 bp in length
2785 2884: gap of 100 bp
2885 4130: contig of 1246 bp in length
4131 4230: gap of 100 bp
4231 6100: contig of 1870 bp in length
6101 6200: gap of 100 bp
6201 7349: contig of 1149 bp in length
7350 7449: gap of 100 bp
7450 9196: contig of 1747 bp in length
9197 9296: gap of 100 bp
9297 10406: contig of 1110 bp in length
10407 10506: gap of 100 bp
10507 12527: contig of 2021 bp in length
12528 12627: gap of 100 bp
12628 14448: contig of 1821 bp in length
14449 14548: gap of 100 bp
14549 16709: contig of 2161 bp in length
16710 16809: gap of 100 bp
16810 18297: contig of 1488 bp in length
18298 18397: gap of 100 bp
18398 20768: contig of 2371 bp in length
20769 20868: gap of 100 bp
20869 23477: contig of 2609 bp in length
23478 23577: gap of 100 bp
23578 26149: contig of 2572 bp in length
26150 26249: gap of 100 bp
26250 30689: contig of 4440 bp in length
30690 30789: gap of 100 bp
30790 37563: contig of 6774 bp in length
37564 37663: gap of 100 bp
37664 47762: contig of 10099 bp in length
47763 47862: gap of 100 bp
47863 55531: contig of 7669 bp in length
55532 55631: gap of 100 bp
55632 64314: contig of 8663 bp in length
64315 64414: gap of 100 bp
64415 75211: contig of 10797 bp in length
75212 75311: gap of 100 bp
75312 87080: contig of 11769 bp in length
87081 87180: gap of 100 bp
87181 124016: contig of 36836 bp in length
124017 124116: gap of 100 bp
124117 140583: contig of 16467 bp in length
140584 140683: gap of 100 bp
140684 160978: contig of 20295 bp in length
160979 161078: gap of 100 bp
161079 174982: contig of 13904 bp in length.
FEATURES
source
1..174982
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-237L12"
/clone.lib="RP23-237L12 Female Mouse BAC"
1..1803
/note="assembly-fragment"
clone_end:SP6
vector_side:left
1904..2784
/note="assembly-fragment"
2885..4130
/note="assembly-fragment"
4231..6100
/note="assembly-fragment"
6201..7349
/note="assembly-fragment"
7450..9196
/note="assembly-fragment"
9297..10406

```

misc_feature      /note="assembly_fragment"
10507. .12527
/note="assembly_fragment"
misc_feature      12628. 14448
/note="assembly_fragment"
misc_feature      14349. 16709
/note="assembly_fragment"
misc_feature      16810. 18297
/note="assembly_fragment"
misc_feature      18398. 20768
/note="assembly_fragment"
misc_feature      20869. 23477
/note="assembly_fragment"
misc_feature      23578. 26149
/note="assembly_fragment"
misc_feature      26250. 30689
/note="assembly_fragment"
misc_feature      30790. 37563
/note="assembly_fragment"
misc_feature      37664. 47762
/note="assembly_fragment"
misc_feature      47863. 55531
/note="assembly_fragment"
misc_feature      55632. 64314
/note="assembly_fragment"
misc_feature      64415. 75211
/note="assembly_fragment"
misc_feature      75312. 87080
/note="assembly_fragment"
misc_feature      87181. 124016
/note="assembly_fragment"

```

```

Query Match      0.9% Score 77: DB 2: Length 174982:
Best Local Similarity 100.0%: Pred. No. 2.4e-30:
Matches 77: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

QY 4638 GCAGCAGATTCTGATTCAGAGCCAGCCTGCTATAGACTGATTCAGAGACCA 4697
|||||
DB 148571 GCAGCAGATTCTGATTCAGAGCCAGCCTGCTATAGACTGATTCAGAGACCA 148512
|||||
QY 4698 GGGCTACACAGAAC 4714
|||||
DB 148511 GGGCTACACAGAAC 148495
|||||

```

```

RESULT 12
AC122433      184854 bp DNA linear HTG 06-JUL-2002
DEFINITION Mus musculus chromosome UNK clone RP24-194G3, WORKING DRAFT
ACCESSION AC122433
VERSION AC122433.2 GI:21700727
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 184854)
McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 184854)
McPherson,J.D. and Waterston,R.H.
REFERENCE Direct Submission
TITLE Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 184854)
McPherson,J.D. and Waterston,R.H.
REFERENCE Direct Submission
TITLE Submitted (06-JUL-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
COMMENT On Jul 6, 2002 this sequence version replaced gi:21105893.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BB0194G03
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: Plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177563 bases at least Q40
Consensus quality: 178348 bases at least Q30
Consensus quality: 179250 bases at least Q20
Insert size: 221000; agarose-fp
Insert size: 184885; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 10.13 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1821: contig of 1821 bp in length
1822      1921: gap of unknown length
1922      5618: contig of 3697 bp in length
5619      5718: gap of unknown length
5719      11721: contig of 6003 bp in length
11722      11821: gap of unknown length
11822      18606: contig of 6785 bp in length
18607      18706: gap of unknown length
18707      23368: contig of 4662 bp in length
23369      23468: gap of unknown length
23469      31153: contig of 7685 bp in length
31154      40066: contig of 8813 bp in length
40067      40166: gap of unknown length
40167      53402: contig of 13236 bp in length
53403      53502: gap of unknown length
53503      69231: contig of 15729 bp in length
69232      69331: gap of unknown length
69332      87512: contig of 18181 bp in length
87513      87612: gap of unknown length
87613      117654: contig of 30042 bp in length
117655      117754: gap of unknown length
117755      184854: contig of 67100 bp in length.
Location/Qualifiers
1. 184854
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-194G3"
1. 1821
/note="assembly_name:Contig10"
1922. 5618
/note="assembly_name:Contig11"
5719. 11721
/note="assembly_name:Contig12"
11822. 18606
/note="assembly_name:Contig13"
18707. 23368
/note="assembly_name:Contig14"
23469. 31153
/note="assembly_name:Contig15"
31254. 40066
/note="assembly_name:Contig16"

```

```
misc_feature      40167..53402
                   /note="assembly_name:Contig17"
misc_feature      53503..69231
                   /note="assembly_name:Contig18"
misc_feature      69332..87512
                   /note="assembly_name:Contig19"
misc_feature      87613..117654
                   /note="assembly_name:Contig20"
misc_feature      117755..184854
                   /note="assembly_name:Contig21"
BASE COUNT      49138 a 41968 c 41001 g 51557 t 1190 others
ORIGIN
Query Match      0.9%; Score 77; DB 2; Length 184854;
Best Local Similarity 100.0%; Pred. No. 2.4e-30;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4643 CAGATTTCGAGTTCAGGCGCCGCTGCTATAGAGTTCAGTTCCAGGACGCGGCT 4702
      |||||||
DB 119686 CAGATTTCGAGTTCAGGCGCCGCTGCTATAGAGTTCAGTTCCAGGACGCGGCT 119745
      |||||||
QY 4703 ACACAGAGAACCCTGT 4719
      |||||||
DB 119746 ACACAGAGAACCCTGT 119762
      |||||||

RESULT 13
AC087869          178032 bp      DNA      linear      HTG 24-FEB-2001
LOCUS             Mus musculus clone RP23-297D13, *** SEQUENCING IN PROGRESS ***, 36
DEFINITION        Mus musculus clone RP23-297D13, *** SEQUENCING IN PROGRESS ***, 36
ACCESSION         AC087869
VERSION           AC087869.1 GI:12658016
KEYWORDS          HTG; HTGS; PHASE1.
SOURCE            Mus musculus.
ORGANISM          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE         1 (bases 1 to 178032)
AUTHORS           McCombie,W.R., Baker,J.P., Bahret,A., Bai,H., Balija,V.,
                  Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
                  Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Yil,M.D.
                  and Zutterer,T.
TITLE            Mouse Genomic Sequence
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 178032)
AUTHORS           McCombie,W.R.
JOURNAL           Direct Submission
TITLE            Submitted (03-FEB-2001) Lita Annenberg Hazen Genome Sequencing
JOURNAL           Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
                  Harbor, NY 11724, USA
COMMENT           ----- Genome Center
                  Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
                  Laboratory
                  Center code: CSHL
                  Web site: http://www.cshl.org/geneseq
                  Contact: mcombie@cshl.org
                  ----- Project Information
                  Center project name: RP23-297D13
                  Center clone name: RP23-297D13
                  -----
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 36 contigs. The true order of the pieces
                  * is not known and their order in this sequence record is
                  * arbitrary. Gaps between the contigs are represented as
                  * runs of N, but the exact sizes of the gaps are unknown.
                  * This record will be updated with the finished sequence
                  * as soon as it is available and the accession number will
                  * be preserved.
                  *
                  * 1 15715: contig of 15715 bp in length
                  * 15716 15757: gap of unknown length
                  * 15758 29656: contig of 13899 bp in length
```

```
* 29657 29698: gap of unknown length
* 29699 42562: contig of 12864 bp in length
* 42563 42604: gap of unknown length
* 42605 52509: contig of 9905 bp in length
* 52510 52551: gap of unknown length
* 52552 61450: contig of 8899 bp in length
* 61451 61492: gap of unknown length
* 61493 68261: contig of 6769 bp in length
* 68262 68303: gap of unknown length
* 68304 74952: contig of 6649 bp in length
* 74953 74994: gap of unknown length
* 74995 81346: contig of 6352 bp in length
* 81347 81388: gap of unknown length
* 81389 87576: contig of 6188 bp in length
* 87577 87618: gap of unknown length
* 87619 93570: contig of 5952 bp in length
* 93571 93612: gap of unknown length
* 93613 99124: contig of 5512 bp in length
* 99125 99166: gap of unknown length
* 99167 104539: contig of 5373 bp in length
* 104540 104581: gap of unknown length
* 104582 109700: contig of 5119 bp in length
* 109701 109741: gap of unknown length
* 109742 114825: contig of 5084 bp in length
* 114826 114867: gap of unknown length
* 114867 119608: contig of 4742 bp in length
* 119609 119649: gap of unknown length
* 119650 123970: contig of 4321 bp in length
* 123971 124011: gap of unknown length
* 124012 128292: contig of 4281 bp in length
* 128293 128333: gap of unknown length
* 128334 132542: contig of 4209 bp in length
* 132543 132583: gap of unknown length
* 132584 136778: contig of 4195 bp in length
* 136779 136819: gap of unknown length
* 136820 140853: contig of 4034 bp in length
* 140854 140894: gap of unknown length
* 140895 144389: contig of 3495 bp in length
* 144390 144390: gap of unknown length
* 144391 147650: contig of 3320 bp in length
* 147651 147691: gap of unknown length
* 147692 150780: contig of 3089 bp in length
* 150781 150821: gap of unknown length
* 150822 153719: contig of 2898 bp in length
* 153720 153760: gap of unknown length
* 153761 156622: contig of 2862 bp in length
* 156623 156663: gap of unknown length
* 156664 159442: contig of 2779 bp in length
* 159443 159483: gap of unknown length
* 159484 161783: contig of 2300 bp in length
* 161784 161824: gap of unknown length
* 161825 164105: contig of 2281 bp in length
* 164106 164146: gap of unknown length
* 164147 166390: contig of 2244 bp in length
* 166391 166431: gap of unknown length
* 166432 168514: contig of 2083 bp in length
* 168515 168555: gap of unknown length
* 168556 170532: contig of 1977 bp in length
* 170533 170573: gap of unknown length
* 170574 172531: contig of 1958 bp in length
* 172532 172572: gap of unknown length
* 172573 174481: contig of 1909 bp in length
* 174482 174522: gap of unknown length
* 174523 176284: contig of 1762 bp in length
* 176285 176325: gap of unknown length
* 176326 177934: contig of 1609 bp in length
* 177935 177975: gap of unknown length
* 177976 178032: contig of 57 bp in length.
*
FEATURES
SOURCE
1. 178032
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="RP23-297D13"
```



```
BASE COUNT      49436 a 38986 c 38427 g 49678 t 1505 others
ORIGIN
Query Match      0.8%; Score 76; DB 2; Length 178032;
Best Local Similarity 100.0%; Pred. No. 8.5e-30;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4638 GCAGCAGATTTCGATTCAGAGCCAGCCCTGCTATAGTAGTTCAGAGTTCAGAGACGCCA 4697
DB 86489 GCAGCAGATTTCGATTCAGAGCCAGCCAGCCCTGCTATAGTAGTTCAGAGTTCAGAGACGCCA 86548
Qy 4698 GGGCTACACAGAGAA 4713
DB 86549 GGGCTACACAGAGAA 86564

RESULT 14
AC074146      201602 bp      DNA      linear      HTG 15-JUL-2000
LOCUS      Mus musculus clone RP23-107G3, WORKING DRAFT SEQUENCE, 38 unordered
DEFINITION      pieces.
ACCESSION      AC074146
VERSION      AC074146.1 GI:9211256
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201602)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 201602)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-JUL-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-23_107G3
-----
Summary Statistics
Consensus quality: 134464 bases at least Q40
Consensus quality: 159739 bases at least Q30
Consensus quality: 170106 bases at least Q20
Estimated insert size: 181300; agarose-fp estimation
Estimated insert size: 197802; sum-of-contigs estimation
Quality coverage: 3.41 in Q20 bases; agarose-fp estimation
Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the contigs
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1675: contig of 1675 bp in length
* 1676 1775: gap of unknown length
* 1776 3291: contig of 1516 bp in length
* 3292 3391: gap of unknown length
* 3392 4432: contig of 1041 bp in length
* 4433 6053: contig of 1521 bp in length
* 6054 6153: gap of unknown length
* 6154 7161: contig of 1008 bp in length
* 7162 7261: gap of unknown length
* 7262 8417: contig of 1156 bp in length
```

```
* 8418 8517: gap of unknown length
* 8518 9800: contig of 1283 bp in length
* 9801 9900: gap of unknown length
* 9901 10926: contig of 1026 bp in length
* 10927 11027: gap of unknown length
* 11027 12062: contig of 1036 bp in length
* 12063 12162: gap of unknown length
* 12163 14110: contig of 1948 bp in length
* 14111 14210: gap of unknown length
* 14211 15534: contig of 1324 bp in length
* 15535 15634: gap of unknown length
* 15635 17316: contig of 1682 bp in length
* 17317 17416: gap of unknown length
* 17417 20982: contig of 3566 bp in length
* 20983 21082: gap of unknown length
* 21083 24936: contig of 3854 bp in length
* 24937 25036: gap of unknown length
* 25037 28124: contig of 3088 bp in length
* 28125 30948: gap of unknown length
* 30949 31048: contig of 2724 bp in length
* 31049 34011: gap of unknown length
* 34012 34111: contig of 2963 bp in length
* 34112 37382: contig of 3371 bp in length
* 37383 37482: gap of unknown length
* 37483 41137: contig of 3654 bp in length
* 41137 41236: gap of unknown length
* 41237 45754: contig of 4518 bp in length
* 45755 45854: gap of unknown length
* 45855 48474: contig of 2620 bp in length
* 48475 48574: gap of unknown length
* 48575 52059: contig of 3485 bp in length
* 52060 52159: gap of unknown length
* 52160 55993: contig of 3634 bp in length
* 55994 56093: gap of unknown length
* 56094 59136: contig of 3043 bp in length
* 59137 59236: gap of unknown length
* 59237 63015: contig of 3779 bp in length
* 63016 63115: gap of unknown length
* 63116 66817: contig of 3802 bp in length
* 66818 67017: gap of unknown length
* 67018 72225: contig of 5208 bp in length
* 72226 72325: gap of unknown length
* 72326 80793: contig of 8468 bp in length
* 80794 80893: gap of unknown length
* 80894 88129: contig of 7236 bp in length
* 88130 88229: gap of unknown length
* 88230 95766: contig of 7537 bp in length
* 95767 95866: gap of unknown length
* 95867 104692: contig of 8826 bp in length
* 104693 104792: gap of unknown length
* 104793 113453: contig of 8661 bp in length
* 113454 113553: gap of unknown length
* 113554 121609: contig of 8056 bp in length
* 121610 121709: gap of unknown length
* 121710 132117: contig of 10408 bp in length
* 132118 132217: gap of unknown length
* 132218 143836: contig of 11619 bp in length
* 143837 143936: gap of unknown length
* 143937 153823: contig of 9887 bp in length
* 153824 153923: gap of unknown length
* 153924 174899: contig of 20976 bp in length
* 174900 174999: gap of unknown length
* 175000 201602: contig of 26603 bp in length.
Location/Qualifiers
1. .201602
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-107G3"
/location="RPCI mouse BAC library 23"
BASE COUNT      52434 a 47973 c 47287 g 49607 t 4301 others
ORIGIN
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:31 : Search time 29.4042 Seconds
(without alignments)
12668.132 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gactggtgactgttgagagagacct 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

| Database : | EST:* |
|--------------------|-------|
| 1: em_estbda:* | |
| 2: em_esthum:* | |
| 3: em_estlin:* | |
| 4: em_estclu:* | |
| 5: em_estcov:* | |
| 6: em_esttpl:* | |
| 7: em_estro:* | |
| 8: em_hic:* | |
| 9: gb_estl:* | |
| 10: gb_estl2:* | |
| 11: gb_hic:* | |
| 12: gb_estl3:* | |
| 13: gb_estl4:* | |
| 14: gb_estl5:* | |
| 15: em_estfun:* | |
| 16: em_estcom:* | |
| 17: gb_gss:* | |
| 18: em_gss_hum:* | |
| 19: em_gss_inv:* | |
| 20: em_gss_pln:* | |
| 21: em_gss_vit:* | |
| 22: em_gss_fun:* | |
| 23: em_gss_mam:* | |
| 24: em_gss_mus:* | |
| 25: em_gss_other:* | |
| 26: em_gss_pro:* | |
| 27: em_gss_rtd:* | |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Score | Query | Length | DB | ID | Description |
|--------|-------|-------|--------|----|----------|-------------------|
| No. | | | | | | |
| 1 | 23 | 100.0 | 175 | 10 | BB583035 | BB583035 BB583035 |
| 2 | 23 | 100.0 | 225 | 10 | BB583295 | BB583295 BB583295 |
| 3 | 23 | 100.0 | 244 | 10 | BB583549 | BB583549 BB583549 |
| 4 | 23 | 100.0 | 235 | 10 | BB582128 | BB582128 BB582128 |
| 5 | 23 | 100.0 | 284 | 10 | BB846624 | BB846624 BB846624 |
| 6 | 23 | 100.0 | 354 | 10 | BB871941 | BB871941 BB871941 |

| | | | | | | | |
|---|----|------|-------|------|----|----------|----------|
| C | 7 | 23 | 100.0 | 414 | 10 | BB848113 | BB848113 |
| C | 8 | 23 | 100.0 | 524 | 10 | AM259173 | AM259173 |
| C | 9 | 23 | 100.0 | 637 | 10 | BB661515 | BB661515 |
| C | 10 | 23 | 100.0 | 674 | 13 | B1143380 | B1143380 |
| C | 11 | 23 | 100.0 | 752 | 9 | A1663456 | A1663456 |
| C | 12 | 23 | 100.0 | 764 | 12 | BF785145 | BF785145 |
| C | 13 | 23 | 100.0 | 856 | 13 | B1100945 | B1100945 |
| C | 14 | 23 | 100.0 | 879 | 13 | B1100201 | B1100201 |
| C | 15 | 23 | 100.0 | 897 | 12 | BF781298 | BF781298 |
| C | 16 | 23 | 100.0 | 916 | 14 | BQ956792 | BQ956792 |
| C | 17 | 23 | 100.0 | 936 | 14 | BQ935016 | BQ935016 |
| C | 18 | 23 | 100.0 | 956 | 12 | BF687321 | BF687321 |
| C | 19 | 23 | 100.0 | 997 | 10 | BB609451 | BB609451 |
| C | 20 | 21.4 | 93.0 | 109 | 9 | AU076562 | AU076562 |
| C | 21 | 21.4 | 93.0 | 300 | 9 | AU098819 | AU098819 |
| C | 22 | 21.4 | 93.0 | 481 | 9 | AA070552 | AA070552 |
| C | 23 | 21.4 | 93.0 | 710 | 13 | B1252311 | B1252311 |
| C | 24 | 21.4 | 93.0 | 723 | 13 | B1761873 | B1761873 |
| C | 25 | 21.4 | 93.0 | 796 | 13 | B116224 | B116224 |
| C | 26 | 21.4 | 93.0 | 875 | 14 | BQ652678 | BQ652678 |
| C | 27 | 21.4 | 93.0 | 903 | 14 | BQ650039 | BQ650039 |
| C | 28 | 21.4 | 93.0 | 935 | 14 | BQ651090 | BQ651090 |
| C | 29 | 21.4 | 93.0 | 952 | 14 | BQ647976 | BQ647976 |
| C | 30 | 21.4 | 93.0 | 969 | 14 | BQ650225 | BQ650225 |
| C | 31 | 21.4 | 93.0 | 1014 | 14 | BM923688 | BM923688 |
| C | 32 | 21.4 | 93.0 | 1206 | 14 | BM923496 | BM923496 |
| C | 33 | 21 | 91.3 | 413 | 10 | BB845511 | BB845511 |
| C | 34 | 19.8 | 86.1 | 230 | 10 | AW336826 | AW336826 |
| C | 35 | 19.8 | 86.1 | 441 | 10 | BB847149 | BB847149 |
| C | 36 | 19.8 | 86.1 | 578 | 10 | AV607319 | AV607319 |
| C | 37 | 18.8 | 81.7 | 726 | 17 | AZ448995 | AZ448995 |
| C | 38 | 18.4 | 80.0 | 443 | 12 | BF516764 | BF516764 |
| C | 39 | 18.4 | 80.0 | 499 | 12 | BG275365 | BG275365 |
| C | 40 | 18.4 | 80.0 | 581 | 17 | AZ710407 | AZ710407 |
| C | 41 | 18.2 | 79.1 | 397 | 17 | AQ116749 | AQ116749 |
| C | 42 | 18.2 | 79.1 | 407 | 10 | AV732812 | AV732812 |
| C | 43 | 18.2 | 79.1 | 470 | 10 | AM823714 | AM823714 |
| C | 44 | 18.2 | 79.1 | 486 | 14 | BQ034151 | BQ034151 |
| C | 45 | 18.2 | 79.1 | 493 | 14 | BQ048109 | BQ048109 |

ALIGNMENTS

| | | | | | |
|------------|--|--|--------------|--------|-----------------|
| RESULT 1 | BB583035 | 175 bp | mRNA | linear | EST 30-NOV-2000 |
| LOCUS | BB583035 | | | | |
| DEFINITION | BB583035 | RIKEN full-length enriched, adult male cecum | Mus musculus | | |
| ACCESSION | BB583035 | | | | |
| VERSION | BB583035.1 | GI:11479579 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | house mouse. | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathia; Muridae; Murinae; Mus. | | | | |
| AUTHORS | 1 (bases 1 to 175) | | | | |
| | Alizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodojima,Y., Imotani,K., Ishii,Y., Itoh,M., Iwawa,M., Kawai,J., Kojima,Y., Kono,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toyota,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y. | | | | |
| TITLE | RIKEN Mouse ESTs (Alizawa,K. et al. 2000) | | | | |
| JOURNAL | Unpublished (2000) | | | | |
| COMMENT | Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan | | | | |

Db 115 GAGTGTGATGTTGAGAGACCT 93

RESULT 3
BB583549/c 235 bp mRNA linear EST 30-NOV-2000
LOCUS BB583549 RIKEN full-length enriched, adult male cecum Mus musculus
DEFINITION CDNA clone 9130405B18 5', mRNA sequence.
ACCESSION BB583549
VERSION BB583549.1 GI:11480093
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
1 (bases 1 to 235)
Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojima, Y.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno
H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,
Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..235
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="9130405B18"
/clone_id="RIKEN full-length enriched, adult male cecum"
/sex="male"
/tissue_type="cecum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Salt. Site_2: BamHI: CDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand CDNA was
primed with a primer [5'
GAGGAGAGAGATCCAGAGCGCTTTTCTTTTCTTTTCTTTVN 3'], CDNA was
prepared by using trehalose and subsequently enriched for full-length by
cap-trapper. CDNA went through one round of normalization

to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAGATTCGATGATTAATTAATTAATCCCCCCCC
3']. CDNA was cloned into the XhoI and BamHI sites.
Vector: a modified Bluescript KS(+) after bulk excision
from Lambda FLX I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

BASE COUNT 73 a 54 c 42 g 66 t
ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 235;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 106 GAGTGTGATGTTGAGAGACCT 84

RESULT 4
BB582128/c 244 bp mRNA linear EST 30-NOV-2000
LOCUS BB582128 RIKEN full-length enriched, adult male colon Mus musculus
DEFINITION CDNA clone 9030202P15 5', mRNA sequence.
ACCESSION BB582128
VERSION BB582128.1 GI:11478672
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
1 (bases 1 to 244)
Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P.,
Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojima, Y.,
Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Kojima, Y., Konno
H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K.,
Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..244
/organism="Mus musculus"
/strain="C57BL/6J"
/db.xref="taxon:10090"
/clone="9030202P15"

| | | | | |
|-------------|---|------------------------|---------------|-------------|
| BASE COUNT | 83 a | 63 c | 53 g | 45 t |
| ORIGIN | BamHI" | | | |
| Query Match | 100.0% | Score 23: | DB 10: | Length 244: |
| Best Local | Similarity | 100.0% | Prod. NO. 23: | |
| Matches 23: | Conservative 0: | Mismatches 0: | Indels 0: | Gaps 0: |
| Oy | 1 | GAGTGGTGATGTTGAGAGACCT | 23 | |
| Db | 107 | GAGTGGTGATGTTGAGAGACCT | 85 | |
| RESULT 5 | | | | |
| LOCUS | BB846624 | 284 bp | mRNA | linear |
| DEFINITION | BB846624 RIKEN full-length enriched, adult male kidney Mus musculus | | | |
| ACCESSION | CDNA clone F530003J06 5', mRNA sequence. | | | |
| VERSION | BB846624 | | | |
| KEYWORDS | BB846624.1 | GI:17084999 | | |
| SOURCE | EST. | | | |
| ORGANISM | house mouse. | | | |
| REFERENCE | Mus musculus | | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 284) | | | |
| | Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawal,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shitara,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akaike,S., Tanaka,T., Tomaru,A., Toya,T., Wachihi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. | | | |
| TITLE | RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001) | | | |
| JOURNAL | Unpublished (2001) | | | |
| COMMENT | Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) | | | |

magi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Matahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsutera
,S., Kawai,U., Okazaki,Y., Muramatsu,M., Inoue,T., Kita,A. and
Hayashizaki,Y.
RIKEN Integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.
Location/Qualifiers
1. .284
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="F530003D06"
/clone_id="RIKEN full-length enriched, adult male kidney"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI. Site_2: SctI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGCGCCGCGCAACTGAGTTTGTGTTTGTGTTTAA 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAGAGAGTCATCCACAGGCTCATTTATTTATTTAAACCCCCCCC 3'].
cDNA was cleaved with XhoI and SctI. "
BASE COUNT 67 a 88 c 62 g 67 t
ORIGIN
Query Match 100.0%; Score 23; DB 10; Length 284;
Best Local Similarity 100.0%; Pctd. No. 24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAGTGTGATGTTTGAGAGCGCT 23
|||||
Db 107 GAGTGTGATGTTTGAGAGCGCT 85
RESULT 6
BB871941/c 354 bp mRNA linear EST 27-NOV-2001
LOCUS BB871941 RIKEN full-length enriched, adult male intestinal mucosa
DEFINITION Mus musculus cDNA clone G630033I20 5', mRNA sequence.
ACCESSION BB871941
VERSION BB871941.1 GI:17118151
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 354)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiroka,T., Hirozane,T., Imotani,K., Ishii,
,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akihira,S., Tanaka,T., Tomaru,A., Toya,T.,
Watanhiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.

hizaki

Location/Qualifiers
1. .354

Query Match

RESULT 7

| | |
|-----------|-------------|
| ACCESSION | BB848113 |
| VERSION | BB848113.1 |
| | GI:17089567 |

| SOURCE ORGANISM | |
|---|--|
| house mouse. | |
| Mus musculus | |
| Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | |
| Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus | |

Carninci, P., Furuno, M., Hanagaki, T.,

hizaki

Location/Qualifiers
1. .414

query match

| | | |
|-----------------------------|---------------------------|---------------|
| Best Local Similarity | 100.0% | Pred. No. 28; |
| Matches 23; Conservative 0; | Mismatches 0; | Indels 0; |
| Gaps 0; | | |
| 0y | 1 GAGTGGTCATCTTTGAGAGACCT | 23 |

DB 107 GAGTGTGATGTTGAGAGACCT 85

RESULT 8
LOCUS AM259173/c 524 bp mRNA linear EST 23-DEC-1999
DEFINITION um89d05.y1 Sugano mouse kidney mklia Mus musculus cDNA clone
IMAGE:2331945 5' similar to gb:U98454 Mus musculus villin protein
mRNA, complete cds (MOUSE);, mRNA sequence.

ACCESSION AM259173 GI:6632154
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 524)
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Riltter, E., Kohn, S., Shio, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The WashU-MCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-MCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the MGI Consortium (info@image.llnl.gov) for further information.
MGI:1011149
Seq primer: custom primer used
High quality sequence stop: 433.
Location/Qualifiers
1. 524
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_1b="IMAGE:2331945"
/clone_1lb="Sugano mouse kidney mklia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII (CACCTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACCTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTCTCTCTTAAGCTGCC and 3' end primer CGACTCGACGTGAGACCA."

BASE COUNT 129 a 139 c 144 g 112 t
ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 524;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACCT 23
|||||
DB 87 GAGTGTGATGTTGAGAGACCT 65

RESULT 9
BB661515/c

LOCUS BB661515 637 bp mRNA linear EST 26-OCT-2001
DEFINITION BB661515 RIKEN full-length enriched, 0 day neonate kidney Mus musculus cDNA clone D630043D01 5', mRNA sequence.

ACCESSION BB661515
VERSION BB661515.1 GI:16495294
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 637)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Koda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES
source
Location/Qualifiers
1. 637
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1b="D630043D01"
/clone_1lb="RIKEN full-length enriched, 0 day neonate kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer (5' GAGAGAGAGCGCGCCGACACTGATTTTCTTTTCTTTTCTTTT 3'), cDNA was prepared by using trehalose thermo-activated reverse

| | |
|------------|---|
| RESULT | 12 |
| LOCUS | Bf785145/c |
| DEFINITION | Bf785145 764 bp mRNA linear EST_12-jan-2001 602108329F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236751 |
| ACCESSION | Bf785145 |
| VERSION | Bf785145 |
| KEYWORDS | Bf785145.1 GI:12090181 |
| SOURCE | EST. |
| | house mouse. |

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 764)
NIH-MGC <http://mgc.ncl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM9845 row: i column: 08
High quality sequence stop: 677.
Location/Qualifiers
1. 764
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. 1"

BASE COUNT 193 a 192 c 236 g 143 t

ORIGIN

Query Match 100.0%; Score 23; DB 12; Length 764;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTGAGAGACCT 23
|||||

Db 88 GAGTGTGATGTGAGAGACCT 66

RESULT 13
B1100945/c 856 bp mRNA linear EST 26-JUN-2001
LOCUS 602886269F1 NCI_CGAP_Kid14 Mus musculus CDNA clone IMAGE:5041625
DEFINITION 5' mRNA sequence.
ACCESSION B1100945
VERSION B1100945.1 GI:14551838
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 856)
NIH-MGC <http://mgc.ncl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1114 row: i column: 18
High quality sequence stop: 617.
Location/Qualifiers
1. 856
/organism="Mus musculus"
/strain="FVB/N"

FEATURES
source

/db_xref="taxon:10090"
/clone="IMAGE:5041625"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. 1"

BASE COUNT 249 a 220 c 249 g 138 t

ORIGIN

Query Match 100.0%; Score 23; DB 13; Length 856;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTGAGAGACCT 23
|||||

Db 75 GAGTGTGATGTGAGAGACCT 53

RESULT 14
B1100201/c 879 bp mRNA linear EST 26-JUN-2001
LOCUS 602885242F1 NCI_CGAP_Kid14 Mus musculus CDNA clone IMAGE:5040346
DEFINITION 5' mRNA sequence.
ACCESSION B1100201
VERSION B1100201.1 GI:14551094
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 879)
NIH-MGC <http://mgc.ncl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1111 row: d column: 11
High quality sequence start: 4
High quality sequence stop: 653.
Location/Qualifiers
1. 879
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5040346"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. 1"

BASE COUNT 227 a 256 c 243 g 153 t

ORIGIN

Query Match 100.0%; Score 23; DB 13; Length 879;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTGAGAGACCT 23
|||||

Db 113 GAGTGTGATGTGAGAGACCT 91

RESULT 15

```

BF781298/c      BF781298      897 bp      mRNA      linear      EST 12-JAN-2001
LOCUS           602106974F1 NCI_CGAP_Ki014 Mus musculus cDNA clone IMAGE:4235106
DEFINITION      5' mRNA sequence.
ACCESSION       BF781298
VERSION         BF781298.1 GI:12086334
KEYWORDS
SOURCE          house mouse.
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1 (bases 1 to 897)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strusberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L1AM9841 row: d column: 19
                High quality sequence stop: 706.
FEATURES
Source          1. 897
                location/Qualifiers
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4235106"
                /clone_lib="NCI_CGAP_Ki014"
                /lab_host="DH10B (TI phage-resistant)"
                /note="Organ: Kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.75 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT      219 a      244 c      277 g      157 t
ORIGIN
Query Match     100.0%; Score 23; DB 12; Length 897;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAGTGGTGTGAGAGACCT 23
    |||||||||||||||||||
Db 91 GAGTGGTGTGAGAGACCT 69

Search completed: January 24, 2003, 18:53:16
Job time : 33.4042 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:33 : Search time 46.6937 Seconds

(without alignments)
12384.474 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gagtggatgtgtgagagagact 23

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

```

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/US095A.COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US095B.COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US095C.COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US095D.COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US095E.COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US095F.COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US095G.COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US095H.COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US095I.COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US095J.COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US095K.COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US095L.COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US095M.COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US095N.COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US095O.COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US095P.COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US095Q.COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US095R.COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095S.COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095T.COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095U.COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095V.COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095W.COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095X.COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US095Y.COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US095Z.COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096A.COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096B.COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096C.COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US096D.COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US096E.COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US096F.COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US096G.COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US096H.COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US096I.COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US096J.COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US096K.COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US096L.COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US096M.COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US096N.COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US096O.COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US096P.COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US096Q.COMB.seq.*

```

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 23 | 100.0 | 23 | US-09-877-935-2 | Sequence 2, Appl1 |
| 2 | 23 | 100.0 | 380 | US-10-060-066-20 | Sequence 210, Ap |
| 3 | 23 | 100.0 | 3049 | US-60-360-207-15646 | Sequence 15646, A |
| 4 | 21.4 | 93.0 | 98 | US-08-242-057-4 | Sequence 4, Appl1 |
| 5 | 21.4 | 93.0 | 98 | US-08-242-057B-4 | Sequence 4, Appl1 |
| 6 | 21.4 | 93.0 | 98 | US-08-467-454A-4 | Sequence 4, Appl1 |
| 7 | 21.4 | 93.0 | 98 | US-08-467-454B-4 | Sequence 4, Appl1 |
| 8 | 21.4 | 93.0 | 417 | US-09-716-473-667 | Sequence 867, App |
| 9 | 21.4 | 93.0 | 1991 | US-08-242-057B-1 | Sequence 1, Appl1 |
| 10 | 21.4 | 93.0 | 1991 | US-08-467-454B-1 | Sequence 1, Appl1 |
| 11 | 21.4 | 93.0 | 3467 | US-09-205-070-4233 | Sequence 4233, Ap |
| 12 | 21.4 | 93.0 | 3467 | US-09-340-623-4233 | Sequence 4233, Ap |
| 13 | 21.4 | 93.0 | 3467 | US-09-898-888-4233 | Sequence 4233, Ap |
| 14 | 21.4 | 93.0 | 3467 | US-09-898-888A-4233 | Sequence 4233, Ap |
| 15 | 21.4 | 93.0 | 3487 | US-09-496-914A-8145 | Sequence 8145, Ap |
| 16 | 21.4 | 93.0 | 3487 | US-09-560-815A-8145 | Sequence 8145, Ap |
| 17 | 21.4 | 93.0 | 4519 | US-09-359-922-2625 | Sequence 2625, Ap |
| 18 | 21.4 | 93.0 | 4519 | US-09-359-922-2625 | Sequence 2625, Ap |
| 19 | 21.4 | 93.0 | 4519 | US-09-919-002-2625 | Sequence 2625, Ap |
| 20 | 21.4 | 93.0 | 5219 | US-60-172-373-10168 | Sequence 10168, A |
| 21 | 21.4 | 93.0 | 5287 | US-60-276-258-12273 | Sequence 12273, A |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```
C 22 21.4 93.0 5680 23 US-09-606-776-4585 Sequence 4585, Ap
C 23 21.4 93.0 5680 25 US-09-652-126-9168 Sequence 9168, Ap
C 24 21.4 93.0 5680 25 US-09-652-915-10128 Sequence 10128, A
C 25 21.4 93.0 5680 25 US-09-652-916-9200 Sequence 9200, Ap
C 26 21.4 93.0 5680 27 US-09-698-010-12819 Sequence 12819, A
C 27 21.4 93.0 5680 28 US-09-716-473-2399 Sequence 2399, Ap
C 28 21.4 93.0 5680 29 US-09-717-350-5084 Sequence 5084, Ap
C 29 21.4 93.0 5680 29 US-09-726-174-4654 Sequence 4654, Ap
C 30 21.4 93.0 5680 29 US-09-726-174-4654 Sequence 4654, Ap
C 31 19.8 86.1 506 69 US-60-252-833-30632 Sequence 30632, A
C 32 19.8 86.1 506 69 US-60-252-833-30632 Sequence 30632, A
C 33 19.8 86.1 506 69 US-60-252-833-30632 Sequence 30632, A
C 34 19.8 86.1 506 69 US-60-252-833-30632 Sequence 30632, A
C 35 18.4 80.0 343 20 US-09-534-854-10048 Sequence 10048, A
C 36 18.4 80.0 407 25 US-09-654-617-336155 Sequence 336155, A
C 37 18.4 80.0 407 27 US-09-684-016-336155 Sequence 336155, A
C 38 18.4 80.0 407 27 US-09-684-016-336155 Sequence 336155, A
C 39 18.4 80.0 407 60 US-60-162-747-3797 Sequence 3797, Ap
C 40 18.4 80.0 591 69 US-60-253-456-18428 Sequence 18428, A
C 41 18.4 80.0 1674 42 US-10-219-999-29253 Sequence 29253, A
C 42 18.4 80.0 1707 33 US-09-865-439A-20985 Sequence 20985, A
C 43 18.4 80.0 2522 33 US-09-867-716-17994 Sequence 17994, A
C 44 18.4 80.0 2522 33 US-09-886-492-17994 Sequence 17994, A
C 45 18.4 80.0 14702 19 US-09-528-237A-2208 Sequence 2208, Ap
```

ALIGNMENTS

RESULT 1

```
US-09-877-935-2
; Sequence 2, Application US/09877935
; GENERAL INFORMATION:
; APPLICANT: Pinto, Daniel
; APPLICANT: Robine, Sylvie
; APPLICANT: Ualser, Frederic
; APPLICANT: Louvard, Daniel
; TITLE OF INVENTION: REGULATORY SEQUENCES OF THE MOUSE VILIN GENE - USE IN TRANSGENES
; FILE REFERENCE: 13294-002001
; CURRENT APPLICATION NUMBER: US/09/877, 935
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: PCT/EP 98/08009
; PRIOR FILING DATE: 1998-12-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 23
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: oligonucleotide
US-09-877-935-2
```

Query Match

Best Local Similarity 100.0%; Score 23; DB 33; Length 23;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 1 GAGTGTGATGTGAGAGCCT 23
```

RESULT 2

```
US-10-060-066-210/C
; Sequence 210, Application US/10060066
; GENERAL INFORMATION:
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NOVEL MUTATED MAMMALIAN CELLS AND
; FILE REFERENCE: LEX-0304-USA
; CURRENT APPLICATION NUMBER: US/10/060, 066
```

```
; CURRENT FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: US 60/265,574
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 380
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(380)
; OTHER INFORMATION: n = A,T,C or G
US-10-060-066-210
```

Query Match

Best Local Similarity 100.0%; Score 23; DB 39; Length 380;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 65 GAGTGTGATGTGAGAGCCT 43
```

RESULT 3

```
US-60-360-207-15646/C
; Sequence 15646, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CLO01321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 15646
; LENGTH: 3049
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-15646
```

Query Match

Best Local Similarity 100.0%; Score 23; DB 80; Length 3049;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Oy 1 GAGTGTGATGTGAGAGCCT 23
Db 91 GAGTGTGATGTGAGAGCCT 69
```

RESULT 4

```
US-08-242-057-4/C
; Sequence 4, Application US/08242057
; GENERAL INFORMATION:
; APPLICANT: Pringault, Eric
; APPLICANT: Robine, Sylvie
; APPLICANT: Huelt, Christian
; APPLICANT: Babinet, Charles
; APPLICANT: Louvard, Daniel
; TITLE OF INVENTION: Villin Gene Promoter Sequence and Its
; TITLE OF INVENTION: Use in Vectors, Transformed Mammalian Cell Lines,
; TITLE OF INVENTION: Transgenic Animals, and Cell Lines Derived from the
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


FILING DATE: 13-MAY-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/604,905
FILING DATE: 29-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, Leslie A.
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 03495.0096-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1991 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-242-057B-1

Query Match 93.0%; Score 21.4; DB 6; Length 1991;
Best Local Similarity 95.7%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23
Db 1974 GGGTGTGATGTGAGAGCCT 1952

US-08-467-454B-1/C
Sequence 1, Application US/08467454B

GENERAL INFORMATION:
APPLICANT: PRINGAULT, Eric
APPLICANT: ROBIN, Sylvie
APPLICANT: HUET, Christian
APPLICANT: BABINET, Charles
APPLICANT: LOUYARD, Daniel
TITLE OF INVENTION: VILLIN GENE PROMOTER SEQUENCE AND ITS
TITLE OF INVENTION: USE IN VECTORS, TRANSFORMED MAMMALIAN CELL LINES,
TITLE OF INVENTION: TRANSGENIC ANIMALS, AND CELL LINES DERIVED FROM THE
TITLE OF INVENTION: ANIMALS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,454B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/242,057
FILING DATE: 13-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/121,454
FILING DATE: 16-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/604,905
FILING DATE: 29-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, Leslie A.
REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 03495.0096-03000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1991 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-467-454B-1

Query Match 93.0%; Score 21.4; DB 8; Length 1991;
Best Local Similarity 95.7%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23
Db 1974 GGGTGTGATGTGAGAGCCT 1952

US-09-205-070-4233/C
Sequence 4233, Application US/09205070

GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
FILE REFERENCE: 20411-748
CURRENT APPLICATION NUMBER: US/09/205,070
CURRENT FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 45207
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4233
LENGTH: 3467
TYPE: DNA
ORGANISM: Homo sapiens
US-09-205-070-4233

Query Match 93.0%; Score 21.4; DB 16; Length 3467;
Best Local Similarity 95.7%; Pred. No. 30;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23
Db 73 GGGTGTGATGTGAGAGCCT 51

US-09-340-623-4233/C
Sequence 4233, Application US/09340623
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
FILE REFERENCE: 20411-748CON1
CURRENT APPLICATION NUMBER: US/09/340,623
CURRENT FILING DATE: 1999-06-28
EARLIER APPLICATION NUMBER: US 09/205,070
EARLIER FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 45207
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4233
LENGTH: 3467
TYPE: DNA
ORGANISM: Homo sapiens
US-09-340-623-4233

Query Match 93.0%; Score 21.4; DB 17; Length 3467;
Best Local Similarity 95.7%; Pred. No. 30;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      1 GAGTGGTGATGTTGAGAGAGCCT 23  
        | |||||  
Db     73 GCGTGGTGATGTTGAGAGAGCCT 51
```

```

RESULT 13
US-09-898-888-4233/c
: Sequence 4233, Application US/09898888
:
: GENERAL INFORMATION:
:   APPLICANT: Hyseq, Inc.
:   TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
:   TITLE OF INVENTION: LIBRARIES
:   FILE REFERENCE: 20411-748CON1
:   CURRENT APPLICATION NUMBER: US/09/898,888
:   CURRENT FILING DATE: 2001-07-03
:   PRIOR APPLICATION NUMBER: 09/340,623
:   PRIOR FILING DATE: 1999-06-28
:   NUMBER OF SEQ ID NOS: 45207
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 4233
:   LENGTH: 3467
:   TYPE: DNA
:   ORGANISM: Homo sapiens
:
US-09-898-888-4233

```

| Query Match | Similarity | 93.0% | Score 21.4 | DB 33 | Length 3467 |
|-------------|------------|-------------------------|--------------|------------|-------------|
| Best Local | Similarity | 95.7% | Pred. No. 30 | | |
| Matches | 22 | Conservative | 0 | Mismatches | 1 |
| | | | | Indels | 0 |
| | | | | Gaps | 0 |
| Oy | 1 | GAGTGTGATGTTTGAGAGAGCCT | 23 | | |
| | | | | | |
| DB | 73 | GCGGTGATGTTTGAGAGAGCCT | 51 | | |

```
US-09-898-888A-4233/c
RESULT 14
US-09-898-888A-4233/c
Sequence 4233, Application US/09898888A
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
FILE REFERENCE: 20411-748CON1
CURRENT FILING DATE: 2001-07-30
PRIORITY APPLICATION NUMBER: US/09/898,888A
PRIOR FILING DATE: 1999-06-28
PRIOR APPLICATION NUMBER: US 09/205,070
NUMBER OF SEQ ID NOS: 45207
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4233
LENGTH: 3467
TYPE: DNA
ORGANISM: Homo sapiens
US-09-898-888A-4233
```

| | | | | |
|-----------------------|----------------|------------------------|----------|-------------|
| Query Match | 93.0% | Score 21.4 | DB 33 | Length 3467 |
| Best Local Similarity | 95.7% | Pred. No. 30 | | |
| Matches 22 | Conservative 0 | Mismatches 1 | Indels 0 | Gaps 0 |
| QY | 1 | GAGTGGTGTGTTGAGAGAGCCT | 23 | |
| | | | | |
| DB | 73 | GGGTGTGATGTTGAGAGAGCCT | 51 | |

```

RESULT:15
US-09-496-914A-8145/c
; Sequence 8145, Application US/09456914A
; GENERAL INFORMATION:
; APPLICANT: Tang, Yuanhua T.
; APPLICANT: Tillinghast, John
; APPLICANT: Sinku, Ankura
; APPLICANT: Liu, Chenghua

```

```

APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Novel Contigs Obtained
FILE OF INVENTION: From Various Libraries
FILE REFERENCE: 787
CURRENT APPLICATION NUMBER: US/09/496,914A
CURRENT FILING DATE: 2000-02-03
PRIORITY APPLICATION NUMBER: US 09/353,690
PRIORITY FILING DATE: 1999-07-14
PRIORITY APPLICATION NUMBER: US 09/034,341
PRIORITY FILING DATE: 1998-02-13
PRIORITY APPLICATION NUMBER: US 09/045,400
PRIORITY FILING DATE: 1998-03-20
PRIORITY APPLICATION NUMBER: US 09/321,214
PRIORITY FILING DATE: 1999-05-26
PRIORITY APPLICATION NUMBER: US 09/131,598
PRIORITY FILING DATE: 1998-08-10
PRIORITY APPLICATION NUMBER: US 09/431,517
PRIORITY FILING DATE: 1999-11-01
PRIORITY APPLICATION NUMBER: US 09/328,351
PRIORITY FILING DATE: 1999-06-04
PRIORITY APPLICATION NUMBER: US 09/332,782
PRIORITY FILING DATE: 1999-06-14
PRIORITY APPLICATION NUMBER: US 09/225,076
PRIORITY FILING DATE: 1999-01-20
PRIORITY APPLICATION NUMBER: US 09/234,611
PRIORITY FILING DATE: 1999-01-22
PRIORITY APPLICATION NUMBER: US 09/346,956
PRIORITY FILING DATE: 1999-07-02
PRIORITY APPLICATION NUMBER: US 09/362,510
PRIORITY FILING DATE: 1999-07-27
PRIORITY APPLICATION NUMBER: US 09/240,371
PRIORITY FILING DATE: 1999-01-29
PRIORITY APPLICATION NUMBER: US 09/248,797
PRIORITY FILING DATE: 1999-02-12
PRIORITY APPLICATION NUMBER: US 09/271,490
PRIORITY FILING DATE: 1999-03-18
PRIORITY APPLICATION NUMBER: US 09/293,972
PRIORITY FILING DATE: 1999-04-15
PRIORITY APPLICATION NUMBER: US 09/274,861
PRIORITY FILING DATE: 1999-03-23
PRIORITY APPLICATION NUMBER: US 60/125,453
PRIORITY FILING DATE: 1999-03-19
PRIORITY APPLICATION NUMBER: US 60/126,605
PRIORITY FILING DATE: 1999-03-26
PRIORITY APPLICATION NUMBER: US 09/306,350
PRIORITY FILING DATE: 1999-05-07
PRIORITY APPLICATION NUMBER: US 09/399,720
PRIORITY FILING DATE: 1999-09-21
PRIORITY APPLICATION NUMBER: US 09/404,284
PRIORITY FILING DATE: 1999-09-21
PRIORITY APPLICATION NUMBER: US 09/465,877
PRIORITY FILING DATE: 1999-12-16
PRIORITY APPLICATION NUMBER: NOT YET ASSIGNED
PRIORITY FILING DATE: 1999-01-19
NUMBER OF SEQ ID NOS: 10410
SOFTWARE: pt_ct_genes Version 1.02
SEQ ID NO 8145
LENGTH: 3487
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44)...(2526)
OTHER INFORMATION: Similar to g1412172 in the genebank database release 114
US-09-496-914A-8145
OTHER INFORMATION: Run with FASTX 3.3f00, default parameters

```

| | | | | |
|--------------------------|-------|-----------------------|--------|--------------|
| Query Match | 93.0% | Score 21.4; | DB 18; | Length 3487; |
| Best Local Similarity | 95.7% | Pred. No. 30; | | |
| Matches 22: Conservative | 0; | Mismatches | 1; | Indels 0; |
| | | | | Gaps 0; |
| 0Y | 1 | GAGTGTGATCTTGAGAGACCT | 23 | |
| | 1 | | | |

Tue Jan 28 08:20:21 2003

us-09-877-935-2.rnrm

Page 7

Db 92 GGGTGGTCATGTTGAGAGAGCCT 70

Search completed: January 25, 2003, 07:47:10
Job time : 48.6937 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:31 : Search time 4.27201 Seconds
(without alignments)
12124.490 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23
Sequence: 1 gagtggatgtgtgagagacgc 23

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N-Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 23 | 100.0 | 23 | 21 | AAA48516 |
| 2 | 23 | 100.0 | 23 | 21 | AAA49254 |
| 3 | 21.4 | 93.0 | 1990 | 13 | AAO25816 |
| 4 | 21.4 | 93.0 | 2702 | 11 | AAQ04681 |
| 5 | 18.4 | 80.0 | 4286 | 17 | AAI18079 |
| 6 | 18.4 | 80.0 | 4286 | 20 | AAI18079 |
| 7 | 18.2 | 79.1 | 1663 | 24 | ABN88181 |
| 8 | 17.4 | 75.7 | 1475 | 19 | AAV00527 |
| 9 | 17.4 | 75.7 | 1475 | 19 | AAV00528 |

| | | | | | | |
|------|------|------|---------|----|-----------|--------------------|
| C 10 | 17.2 | 74.8 | 538 | 23 | ABV54721 | Human prostate exp |
| C 11 | 17.2 | 74.8 | 963 | 21 | AAC46212 | Arabidopsis thalia |
| C 12 | 17.2 | 74.8 | 1054 | 21 | AAC45514 | Arabidopsis thalia |
| C 13 | 17.2 | 74.8 | 1056 | 21 | AAC37611 | Arabidopsis thalia |
| C 14 | 17.2 | 74.8 | 2884 | 21 | AAI79490 | Eucalyptus grandis |
| C 15 | 17.2 | 74.8 | 3600 | 21 | AAI79657 | Eucalyptus grandis |
| C 16 | 17.2 | 74.8 | 21423 | 22 | AAI36470 | Human musculoskele |
| C 17 | 17.2 | 74.8 | 22073 | 22 | AAI04663 | Human reproductive |
| C 18 | 17.2 | 74.8 | 22073 | 23 | ABL97570 | Human testicular a |
| C 19 | 16.8 | 73.0 | 4774 | 22 | AAK64331 | Human immune/haema |
| C 20 | 16.8 | 73.0 | 2955 | 24 | ABL91678 | Human polynucleoti |
| C 21 | 16.8 | 73.0 | 4281 | 14 | AAO53471 | elk cDNA. Rattus |
| C 22 | 16.8 | 73.0 | 7924 | 24 | ABK40069 | Human chemically p |
| C 23 | 16.8 | 73.0 | 7924 | 24 | ABL34138 | Human immune syste |
| C 24 | 16.8 | 73.0 | 15650 | 22 | ABAI15903 | Human nervous syst |
| C 25 | 16.8 | 73.0 | 15650 | 22 | ABAI16564 | Human nervous syst |
| C 26 | 16.8 | 73.0 | 143068 | 21 | AAE21105 | Human low adenosin |
| C 27 | 16.8 | 73.0 | 143068 | 21 | AAE21272 | Human low adenosin |
| C 28 | 16.8 | 73.0 | 143068 | 21 | AAA34983 | Human adenosine re |
| C 29 | 16.8 | 73.0 | 143068 | 21 | AAA35150 | Human adenosine re |
| C 30 | 16.8 | 73.0 | 143068 | 24 | ABL68124 | Human cancer relat |
| C 31 | 16.8 | 73.0 | 149412 | 21 | AAA35151 | Human adenosine re |
| C 32 | 16.8 | 73.0 | 152740 | 21 | AAE21273 | Human low adenosin |
| C 33 | 16.8 | 73.0 | 2365589 | 24 | ABA90521 | Genomic sequence o |
| C 34 | 16.6 | 72.2 | 321 | 22 | AAK87824 | Human digestive sy |
| C 35 | 16.6 | 72.2 | 387 | 24 | ABN24942 | Human ORF polynuc |
| C 36 | 16.6 | 72.2 | 600 | 21 | AAE08568 | Fusarium venenatum |
| C 37 | 16.6 | 72.2 | 602 | 22 | AAH87718 | Peppermint plant o |
| C 38 | 16.6 | 72.2 | 4618 | 23 | ABK42604 | Genomic sequence # |
| C 39 | 16.6 | 72.2 | 7353 | 24 | ABN89472 | Human G-protein co |
| C 40 | 16.6 | 72.2 | 12989 | 23 | ABK42602 | Genomic sequence # |
| C 41 | 16.6 | 72.2 | 24978 | 20 | AAK60209 | SEQ ID 3 of US9314 |
| C 42 | 16.6 | 72.2 | 24979 | 21 | AAA52321 | Genomic DNA sequen |
| C 43 | 16.6 | 72.2 | 129722 | 24 | ABQ88117 | Human osteoblast d |
| C 44 | 16.4 | 71.3 | 904 | 22 | AAH32619 | Human secreted pro |
| C 45 | 16.4 | 71.3 | 975 | 23 | AA594424 | DNA encoding novel |

ALIGNMENTS

RESULT 1
AAA48516
ID AAA48516 standard, DNA; 23 BP.

XX
AC AAA48516;
XX
XX
DT 19-DEC-2000 (first entry)
XX
DE Murine villin gene regulatory region probe #1.
XX
KW Mouse; villin; regulatory region; digestive tract;
KW colorectal cancer mouse model; probe; ss.
XX
OS Mus sp.
XX
PN WO200034492-A1.
XX
PD 15-JUN-2000.
XX
PF 09-DEC-1998; 98WO-EP08009.
XX
PR 09-DEC-1998; 98WO-EP08009.
XX
PA (CNRS) CENT NAT RECH SCI.
PA (CURT-) INST CURIE.
XX
PI Pinto D, Robine S, Jaisser F, Louvard D;
XX
DR WPI: 2000-423433/36.
XX
PT Novel nucleotide sequence derived from mouse villin gene for targeted expression of transgenes in immature and differentiated epithelial

PT cells of intestine or urogenital tracts -
XX
PS Disclosure; Page 13; 54pp; English.
XX
CC The present sequence is a probe which was used in primer extension
CC analysis of the murine villin gene regulatory region. It has been
CC shown that this region directs the expression of the villin gene in the
CC intestine and uro-genital tracts, and thus could be used in a fusion
CC gene to direct expression of exogenous genes in these areas. This could
CC be used, for example, to create a mouse model for colorectal cancer.
XX
SQ Sequence 23 BP; 5 A; 2 C; 10 G; 6 T; 0 other;
Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAGTGTGATGTTGAGAGAGCCT 23
DB 1 GAGTGTGATGTTGAGAGAGCCT 23
RESULT 2
ID AAA49254 standard; DNA: 23 BP.
XX
AC AAA49254;
XX
DT 19-DEC-2000 (first entry)
XX
DE Mouse villin gene oligonucleotide probe #1.
XX
KM Mouse; villin; intestinal epithelial cell;
KM uro-genital tract epithelial cell; tumour; probe; ss.
XX
OS Mus sp.
XX
PN WO200034493-A2.
XX
PD 15-JUN-2000.
XX
PE 09-DEC-1999; 99WO-EP09782.
XX
PR 09-DEC-1998; 98WO-EP08009.
XX
PA (CNRS) CENT NAT RECH SCI.
PA (CURT-) INST CURIE.
XX
PI Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J;
DR WPI; 2000-423434/36.
XX
PT Novel nucleotide sequence derived from mouse villin gene for targeted
PT expression of transgenes in immature and differentiated epithelial
PT cells of intestine or urogenital tracts -
XX
PS Example 1; Page 13; 52pp; English.
XX
CC The present sequence is an oligonucleotide probe for the murine
CC villin gene. This gene is expressed in the epithelial cells of the
CC intestine and uro-genital tracts. Its promoter sequence can be used in
CC the targeted expression of exogenous genes in these places, which may,
CC for example, be useful in the treatment of tumours. This probe was used
CC in a primer extension assay.
XX
SQ Sequence 23 BP; 5 A; 2 C; 10 G; 6 T; 0 other;
Query Match 100.0%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAGTGTGATGTTGAGAGAGCCT 23
DB 1 GAGTGTGATGTTGAGAGAGCCT 23

DB 1 GAGTGTGATGTTGAGAGAGCCT 23
RESULT 3
ID AAQ25816/c
XX AAQ25816 standard; DNA: 1990 BP.
XX
AC AAQ25816;
XX
DT 06-JAN-1993 (first entry)
XX
DE Villin gene promoter.
XX
KM Oncogene; embryogenesis; terminal differentiation; cell lineage;
KM carcinogenesis; metastasis; chloramphenicol acetyltransferase; CAT;
KM human growth hormone; hGH; luciferase; beta-galactosidase; ds.
XX
OS Homo sapiens.
XX
PN CA2054149-A.
XX
PD 30-APR-1992.
XX
PE 24-OCT-1991; 91CA-2054149.
XX
PR 29-OCT-1990; 90US-0604905.
XX
PA (PRIN/) PRINGAUT E.
XX
PI Babinet C, Huët C, Louvard D, Pringault E, Robine S, Louvard C;
XX
DR WPI; 1992-235048/29.
XX
PT Promoter sequence from human villin gene - used in vectors linked
PT to e.g. oncogene for prodn. of transgenic animal models of
PT colorectal cancer
XX
PS Disclosure; Fig 4"; 63pp; English.
XX
CC This promoter sequence can be operably linked to a gene eg. a an
CC oncogene. Vectors containing such constructs are useful in the
CC production of animal models of eg. colorectal cancer. Villin gene
CC expression is strictly regulated during embryogenesis and terminal
CC differentiation and villin is a marker of digestive epithelial cell
CC lineage. The tissue-specific expression is maintained during
CC carcinogenesis and villin can therefore be used to characterise the
CC colorectal origin of unknown metastasis. The promoter can also be
CC linked to a structural gene and used in transient assay systems based
CC on the use of fusion genes. Suitable reporter genes are
CC chloramphenicol acetyltransferase (CAT), human growth hormone (hGH),
CC luciferase and beta-galactosidase. This promoter is also useful in
CC the investigation of the molecular genetics involved in digestive
CC organogenesis and may be used as a probe for the detection of
CC a nucleotide sequence in a biological material.
XX
SQ Sequence 1990 BP; 549 A; 505 C; 546 G; 390 T; 0 other;
Query Match 93.0%; Score 21.4; DB 13; Length 1990;
Best Local Similarity 95.7%; Pred. No. 2.4;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GAGTGTGATGTTGAGAGAGCCT 23
DB 1973 GGTGTGATGTTGAGAGAGCCT 1951
RESULT 4
ID AAQ04681/c
XX AAQ04681 standard; DNA: 2702 BP.
XX
AC AAQ04681;
XX
DT 20-SEP-1990 (first entry)

```

XX DE Human villin gene fragments for probes and primers.
XX XX
XX KW Human villin gene; amplification primers; probes; tumour diagnosis;
XX KW RFLP; ss
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 25..2505
XX FT /*tag= a
XX FT /product= human villin
XX PN EP367666-A.
XX PD 09-MAY-1990.
XX PF 27-OCT-1989; 89EP-0402984.
XX PR 28-OCT-1988; 88FR-0014208; EP-402984.
XX PA (INSP ) INST PASTEUR.
XX P1 Arpin M, Pringault E, Garcia A, Louvard D;
XX DR WPI: 1990-141549/19.
XX DR P-PSDB: AAR05049.
XX PT New nucleic acid fragments corresponding to villin gene - useful as
XX PT diagnostic amplification primers or hybridisation probes.
XX PS Disclosure: ; P: French.
XX XX
XX CC Any nucleic acid fragment comprising a sequence of 8-40 ,pref.20-40,
XX CC nucleotides which is present in the human villin gene, or is exactly
XX CC complementary to such a sequence, or is present in RNA fragment corresp.
XX CC to such a sequence is useful as primer for amplifying villin-encoding DNA
XX CC or the corresp. mRNA in biological samples, esp. in the diagnosis of
XX CC gastrointestinal tumours or metastases, and as probes for detecting
XX CC RFLPs of the human genome.
XX SQ Sequence 2702 BP; 681 A; 719 C; 778 G; 524 T; 0 other;
XX
XX Query Match 93.0%; Score 21.4; DB 11; Length 2702;
XX Best Local Similarity 95.7%; Pred. No. 2.5;
XX Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGTGTGATGTTGAGAGAGCCT 23
DB 73 GGGTGGTGTGTTGAGAGAGCCT 51

```

```

XX XX
XX PR 27-MAY-1994; 94JP-0115597.
XX XX
XX PA (NASC-) NAT SCI COUNCIL.
XX XX
XX DR WPI: 1996-154854/16.
XX XX
XX PT Prevention of virus infections by induction of apoptosis - uses
XX PT anti-virus gene hem1 (persistence-associated gene 1) to cause
XX PT programmed cell death upon infection
XX PS Disclosure: Page 16-17; 28pp; Japanese.
XX XX
XX CC This sequence represents the HZ-1 baculovirus hem1 gene. Hem1 is also
XX CC known as the persistence-associated gene 1. This sequence can be used in
XX CC the method of the invention for prevention and/or treatment of viral
XX CC infections, by induction of earlier apoptosis in the animal host cells or
XX CC living body. This method comprises making the animal host cells
XX CC susceptible to apoptosis induction upon virus infection and then inducing
XX CC apoptosis by attack of the infectious virus on the host animal cells.
XX CC This causes cell death, followed by blocking the regeneration of the
XX CC infectious virus, thereby killing the virus in the host animal cells. In
XX CC order to carry out this method, the host cells contain an anti-virus gene
XX CC (such as this gene). This gene is preferably introduced to the cell by
XX CC genetic recombination. This method may be used for the prevention of
XX CC virus infection.
XX SQ Sequence 4286 BP; 1256 A; 791 C; 882 G; 1357 T; 0 other;
XX
XX Query Match 80.0%; Score 18.4; DB 17; Length 4286;
XX Best Local Similarity 95.0%; Pred. No. 66;
XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGTGTGATGTTGAGAGAG 20
DB 144 GAGGGGTGATGTTGAGAGAG 125

```

```

RESULT 5
AA18079/C
ID AA18079 standard; DNA; 4286 BP.
XX
XX AC AA18079;
XX
XX 14-AUG-1996 (first entry)
XX
XX Hem1 gene.
XX
XX Hem1; persistence associated gene 1; HZ-1 baculovirus; prevention;
XX KW therapy; viral infection; apoptosis induction; cell death;
XX KW anti-virus gene; ss.
XX OS HZ-1 baculovirus.
XX
XX PN JP08038181-A.
XX
XX 13-FEB-1996.
XX
XX 10-JUN-1994; 94JP-0129283.

```

```

RESULT 6
AA176880/C
ID AA176880 standard; DNA; 4286 BP.
XX
XX AC AA176880;
XX
XX 05-AUG-1999 (first entry)
XX
XX DE HZ-1 pag1 promoter.
XX
XX HZ-1 pag1 promoter; persistence-associated gene 1; insect cell;
XX KW constitutive expression promoter; ss.
XX
XX OS HZ-1 virus.
XX
XX PN US5911982-A.
XX
XX 15-JUN-1999.
XX
XX 18-APR-1996; 96US-0634350.
XX
XX 18-APR-1996; 96US-0634350.
XX PR 06-OCT-1995; 95US-0004894.
XX PR 11-OCT-1995; 95US-0005128.
XX
XX (NASC-) NAT SCI COUNCIL.
XX
XX Chao Y;
XX
XX WPI: 1999-357167/30.
XX
XX HZ-1 virus persistence-associated gene promoter
XX
XX Claim 1; Fig 1b; 56pp; English.
XX

```

CC This sequence represents the Hx-1 persistence-associated gene 1
 CC (Hx-1 pag1) promoter of the invention. The pag1 gene promoter is useful
 CC in insect cells for driving constitutive expression of e.g. genes
 CC encoding foreign proteins. The promoter of the pag1 gene is
 CC constitutively expressed and stronger than that of the polyhedrin gene in
 CC insect cells, enabling it to express foreign genes more strongly
 CC e.g. lacZ and luciferase, in addition to which it can be expressed more
 CC prominently as a short promoter.

XX
 SQ Sequence 4286 BP; 1256 A; 792 C; 881 G; 1357 T; 0 other;

Query Match 80.0%; Score 18.4; DB 20; Length 4286;
 Best Local Similarity 95.0%; Pred. No. 66;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAG 20
 ||| ||||| ||||| |||||
 DB 144 GAGGCGTGTGTGAGAG 125

RESULT 7

ID ABN8181 standard; DNA: 1663 BP.

AC ABN8181;

DT 13-AUG-2002 (first entry)

DE G protein-coupled receptor CGR95 nucleotide sequence.

XX RNA analysis: identification; RNA molecule: antibacterial; virucide;

KW fungicide; cytostatic; antisense therapy; cancer; infection;

KM G protein-coupled receptor; gene; ds.

XX Unidentified.

PN WO200224950-A2.

PD 28-MAR-2002.

PF 25-SEP-2001; 2001WO-SE02054.

PR 25-SEP-2000; 2000US-235029P.

PA (NEUR-) NEUROMICS INC.

PI Liang Z, Zhang H, Wahlestedt C;

DR WPI; 2002-404959/43.

PT Identifying accessible region (AR) of native RNA. Involves selecting
 PT from oligonucleotide population, an oligonucleotide binding to AR,
 PT sequencing randomized portion of oligonucleotide, and identifying
 PT sequence of AR -

XX Example: Fig 8; 41pp; English.

XX The present invention describes a method (M1) for the single-cycle
 CC identification of an accessible region (AR) of native RNA (I). The method
 CC comprises providing an in vitro reaction mixture comprising (i) and a
 CC population of oligonucleotides (II), each having a randomized portion
 CC that can bind to a complementary AR of (I), if present, selecting a (II)
 CC that binds to an AR, sequencing the randomized portion of (II), and
 CC identifying the nucleotide sequence of the AR. An AR can have virucide,
 CC antibacterial, fungicide and cytostatic activities, and can be used in
 CC antineoplastic therapy. The method of the invention is used for identifying
 CC an AR of a native RNA preferably mRNA. Identifying an AR of mRNA is
 CC useful for manufacturing an antisense oligonucleotide for the
 CC downregulation of expression of an mRNA molecule which involves
 CC identifying an AR on an mRNA using the method and synthesizing an
 CC oligonucleotide complementary to AR. (M1) is useful for making an
 CC antisense oligonucleotide which involves identifying an AR of a native
 CC RNA by (M1) and synthesizing the antisense oligonucleotide that is

CC complementary to the AR. The antisense oligonucleotides are useful for
 CC treating disorders associated with aberrant gene expression, such as
 CC cancer and disorders associated with expression of foreign genes such
 CC as infection with bacterial, viral or fungal pathogen. The present
 CC sequence represents an oligonucleotide which is used in the
 CC exemplification of the present invention.

XX
 SQ Sequence 1663 BP; 341 A; 461 C; 440 G; 421 T; 0 other;

Query Match 79.1%; Score 18.2; DB 24; Length 1663;
 Best Local Similarity 87.0%; Pred. No. 71;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTGAGAGCCT 23
 ||| ||||| ||| |||||
 DB 106 GAGTGTGATGTAAAGCAGCCT 84

RESULT 8

ID AAV00527 standard; DNA: 1475 BP.

AC AAV00527;

DT 20-JUL-1998 (first entry)

DE HIV-1 breakthrough isolate clone C10.5 gp120 polypeptide DNA.

XX HIV-1; envelope protein; gp120; MN-rp120; vaccine; AIDS; ss.

KW Human Immunodeficiency virus type 1.

OS Key Location/Qualifiers

FT CDS 2..1450

FT /tag= a

PN WO9801564-A1.

PD 15-JAN-1998.

PF 03-JUL-1997; 97WO-US09690.

PR 08-JUL-1996; 96US-0676737.

PA (GETH) GENENTECH INC.

PI Berman PW;

DR WPI; 1998-101059/09.

DR P-PSDB; AAM37064.

PT Human immunodeficiency virus gp120 sequences from vaccine
 PT breakthrough strains - useful in providing added protection against
 PT HIV not provided by original vaccine

XX Claim 6: Page 133-135; 193pp; English.

XX Clone C10.5 encodes a gp120 envelope polypeptide (see AAM37064) of
 CC a breakthrough isolate of HIV-1 obtained from an individual
 CC infected with HIV-1 through high risk activity while participating
 CC in Phase I or Phase 2 trials of MN-rp120, a candidate recombinant
 CC gp120 HIV-1 vaccine. Nucleotide sequences (see AAV00517-30) for
 CC gp120 polypeptides (see AAM37054-67) were obtained from 2 clones of
 CC each of 7 breakthrough isolates. The envelope glycoprotein genes
 CC were obtained from proviral DNA using PCR (see AAV00533-38). All 7
 CC envelope glycoproteins possess sequences typical of subtype (clade)
 CC B viruses. The overall homology with MN-rp120 is 69-80%. Use of
 CC the gp120 polypeptides from one or more of the breakthrough
 CC isolates, usually together with MN-rp120, can provide protection
 CC against HIV strains that are sufficiently different from the
 CC vaccine strain (e.g. MN-rp120) that the vaccine does not confer
 CC protection against those strains.

SO Sequence 1475 BP; 568 A; 240 C; 314 G; 353 T; 0 other;

Query Match 75.7%; Score 17.4; DB 19; Length 1475;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GTGGTGATGTTGAGAGAC 21
|||||
DB 375 GTGGTGATGTTGAGAGAC 357

RESULT 9
AAV00528/C
ID AAV00528 standard; DNA; 1475 BP.

AC AAV00528;

DT 20-JUL-1998 (first entry)

DE HIV-1 breakthrough isolate clone C10.7 gp120 polypeptide DNA.

XX HIV-1; envelope protein; gp120; MN-rgp120; vaccine; AIDS; ss.

OS Human immunodeficiency virus type 1.

XX Key Location/Qualifiers

FT CDS 2..1465
/*tag= a

PN MO9801564-A1.

PD 15-JAN-1998.

XX 03-JUL-1997; 97WO-US09690.

PR 08-JUL-1996; 96US-0676737.

PA (GENE) GENENTECH INC.

PI Berman PW;

DR WPI; 1998-101059/09.

DR P-PSDB; AAM37065.

PT Human immunodeficiency virus gp120 sequences from vaccine
breakthrough strains - useful in providing added protection against
HIV not provided by original vaccine

PS Claim 6; Page 137-139; 193pp; English.

XX Clone C10.7 encodes a gp120 envelope polypeptide (see AAM37065) of
a breakthrough isolate of HIV-1 obtained from an individual
infected with HIV-1 through high risk activity while participating
in Phase I or Phase 2 trials of MN-rgp120, a candidate recombinant
gp120 HIV-1 vaccine. Nucleotide sequences (see AAV00517-30) for
gp120 polypeptides (see AAM37054-67) were obtained from 2 clones of
each of 7 breakthrough isolates. The envelope glycoprotein genes
were obtained from proviral DNA using PCR (see AAV00533-38). All 7
envelope glycoproteins possess sequences typical of subtype (clade)
B viruses. The overall homology with MN-rgp120 is 69-80%. Use of
the gp120 polypeptides from one or more of the breakthrough
isolates, usually together with MN-rgp120, can provide protection
against HIV strains that are sufficiently different from the
vaccine strain (e.g. MN-rgp120) that the vaccine does not confer
protection against those strains.

SO Sequence 1475 BP; 566 A; 240 C; 316 G; 353 T; 0 other;

Query Match 75.7%; Score 17.4; DB 19; Length 1475;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GTGGTGATGTTGAGAGAC 21

DB 375 GTGGTGATGTTGAGAGAC 357
|||||

RESULT 10
ABV54721/C
ID ABV54721 standard; cDNA; 538 BP.

AC ABV54721;

DT 17-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 54712.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

XX Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-183119P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

PS Claim 1; Page 10575; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification, or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

SO Sequence 538 BP; 113 A; 134 C; 139 G; 152 T; 0 other;

Query Match 74.8%; Score 17.2; DB 23; Length 538;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 ACTGCTGATGTTGAGAGACCT 23
|||||

DB 314 ACTGCTGATGTTGAGAGACCT 293

RESULT 11
AAC46212/C

| | | | | |
|----|--|----|--------------|---------------|
| ID | AAC46212 standard; DNA; 963 BP. | PR | 18-JUN-1999; | 99US-0139461. |
| XX | | PR | 18-JUN-1999; | 99US-0139462. |
| AC | AAC46212; | PR | 18-JUN-1999; | 99US-0139463. |
| XX | | PR | 18-JUN-1999; | 99US-0139464. |
| DT | 18-OCT-2000 (first entry) | PR | 18-JUN-1999; | 99US-0139465. |
| XX | | PR | 18-JUN-1999; | 99US-0139466. |
| DE | Arabidopsis thaliana DNA fragment SEQ ID NO: 49311. | PR | 21-JUN-1999; | 99US-0139467. |
| XX | | PR | 22-JUN-1999; | 99US-0139468. |
| XX | | PR | 23-JUN-1999; | 99US-0140353. |
| KM | Hybridisation assay; genetic mapping; gene expression control; | PR | 23-JUN-1999; | 99US-0140354. |
| KW | Protein identification; signal transduction pathway; | PR | 24-JUN-1999; | 99US-0140695. |
| KM | metabolic pathway; promoter; termination sequence; ss. | PR | 28-JUN-1999; | 99US-0140823. |
| XX | | PR | 29-JUN-1999; | 99US-0140891. |
| OS | Arabidopsis thaliana. | PR | 30-JUN-1999; | 99US-0141287. |
| XX | | PR | 01-JUL-1999; | 99US-0141842. |
| PN | EP1033405-A2. | PR | 01-JUL-1999; | 99US-0142154. |
| XX | | PR | 02-JUL-1999; | 99US-0142055. |
| PD | 06-SEP-2000. | PR | 06-JUL-1999; | 99US-0142390. |
| XX | | PR | 08-JUL-1999; | 99US-0142803. |
| PF | 25-FEB-2000; 2000EP-0301439. | PR | 09-JUL-1999; | 99US-0142920. |
| XX | | PR | 12-JUL-1999; | 99US-0142977. |
| XX | | PR | 13-JUL-1999; | 99US-0143542. |
| PR | 25-FEB-1999; | PR | 14-JUL-1999; | 99US-0143624. |
| PR | 05-MAR-1999; | PR | 15-JUL-1999; | 99US-0144005. |
| PR | 09-MAR-1999; | PR | 16-JUL-1999; | 99US-0144085. |
| PR | 23-MAR-1999; | PR | 16-JUL-1999; | 99US-0144086. |
| PR | 25-MAR-1999; | PR | 19-JUL-1999; | 99US-0144325. |
| PR | 29-MAR-1999; | PR | 19-JUL-1999; | 99US-0144331. |
| PR | 01-APR-1999; | PR | 19-JUL-1999; | 99US-0144332. |
| PR | 06-APR-1999; | PR | 19-JUL-1999; | 99US-0144333. |
| PR | 08-APR-1999; | PR | 19-JUL-1999; | 99US-0144334. |
| PR | 16-APR-1999; | PR | 20-JUL-1999; | 99US-0144352. |
| PR | 19-APR-1999; | PR | 20-JUL-1999; | 99US-0144632. |
| PR | 21-APR-1999; | PR | 20-JUL-1999; | 99US-0144884. |
| PR | 23-APR-1999; | PR | 21-JUL-1999; | 99US-0144814. |
| PR | 28-APR-1999; | PR | 21-JUL-1999; | 99US-0145086. |
| PR | 30-APR-1999; | PR | 21-JUL-1999; | 99US-0145088. |
| PR | 04-MAY-1999; | PR | 22-JUL-1999; | 99US-0145085. |
| PR | 05-MAY-1999; | PR | 22-JUL-1999; | 99US-0145087. |
| PR | 06-MAY-1999; | PR | 22-JUL-1999; | 99US-0145089. |
| PR | 07-MAY-1999; | PR | 22-JUL-1999; | 99US-0145192. |
| PR | 07-MAY-1999; | PR | 23-JUL-1999; | 99US-0145145. |
| PR | 11-MAY-1999; | PR | 23-JUL-1999; | 99US-0145218. |
| PR | 14-MAY-1999; | PR | 23-JUL-1999; | 99US-0145224. |
| PR | 14-MAY-1999; | PR | 26-JUL-1999; | 99US-0145276. |
| PR | 14-MAY-1999; | PR | 27-JUL-1999; | 99US-0145913. |
| PR | 14-MAY-1999; | PR | 27-JUL-1999; | 99US-0145918. |
| PR | 18-MAY-1999; | PR | 27-JUL-1999; | 99US-0145919. |
| PR | 19-MAY-1999; | PR | 28-JUL-1999; | 99US-0145951. |
| PR | 20-MAY-1999; | PR | 02-AUG-1999; | 99US-0146386. |
| PR | 21-MAY-1999; | PR | 02-AUG-1999; | 99US-0146388. |
| PR | 24-MAY-1999; | PR | 02-AUG-1999; | 99US-0146389. |
| PR | 25-MAY-1999; | PR | 03-AUG-1999; | 99US-0147038. |
| PR | 27-MAY-1999; | PR | 04-AUG-1999; | 99US-0147204. |
| PR | 28-MAY-1999; | PR | 05-AUG-1999; | 99US-0147302. |
| PR | 01-JUN-1999; | PR | 05-AUG-1999; | 99US-0147192. |
| PR | 03-JUN-1999; | PR | 05-AUG-1999; | 99US-0147260. |
| PR | 04-JUN-1999; | PR | 06-AUG-1999; | 99US-0147303. |
| PR | 07-JUN-1999; | PR | 06-AUG-1999; | 99US-0147416. |
| PR | 08-JUN-1999; | PR | 09-AUG-1999; | 99US-0147493. |
| PR | 10-JUN-1999; | PR | 09-AUG-1999; | 99US-0147935. |
| PR | 14-JUN-1999; | PR | 10-AUG-1999; | 99US-0148171. |
| PR | 16-JUN-1999; | PR | 11-AUG-1999; | 99US-0148319. |
| PR | 16-JUN-1999; | PR | 12-AUG-1999; | 99US-0148341. |
| PR | 17-JUN-1999; | PR | 13-AUG-1999; | 99US-0148565. |
| PR | 18-JUN-1999; | PR | 13-AUG-1999; | 99US-0148684. |
| PR | 18-JUN-1999; | PR | 16-AUG-1999; | 99US-0149368. |
| PR | 18-JUN-1999; | PR | 17-AUG-1999; | 99US-0149175. |
| PR | 18-JUN-1999; | PR | 18-AUG-1999; | 99US-0149426. |
| PR | 18-JUN-1999; | PR | 20-AUG-1999; | 99US-0149722. |
| PR | 18-JUN-1999; | PR | 20-AUG-1999; | 99US-0149723. |
| PR | 18-JUN-1999; | PR | 23-AUG-1999; | 99US-0149929. |
| PR | 18-JUN-1999; | PR | 23-AUG-1999; | 99US-0149902. |

```
PR 23-AUG-1999; 99US-014930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 74.8%; Score 17.2; DB 21; Length 963;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 GAGTGTGATGTTGAGAGACC 22
   |||||
DB 89 GAGTGTGATGTTGAGAGACC 68
```

RESULT 12
AAC4514/c
ID AAC4514 standard; DNA; 1054 BP.

XX AAC4514;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 46771.

```
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX EPI03405-A2.
XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```

PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 21-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145921.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149829.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.

PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158032.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1054;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGTGGTATGTCAGAGAGCC 22
|||||||
Db 152 GAGTGGTATGTCAGAGAGTCAC 131

RESULT 13
AAC37611/C
ID AAC37611 standard; DNA; 1056 BP.

AC AAC37611;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 18019.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

[illegible]

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157665.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158309.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1056;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 GAGGTGATGTGAGAGACC 22
Db 154 GAGTGTGATGTGAGAGTGC 133

RESULT 14
AAAT9490
ID AAA79490 standard; cDNA; 2884 BP.

AC AAA79490;
DT 27-NOV-2000 (first entry)
DE Eucalyptus grandis cell signalling involved polynucleotide SEQ ID NO:291.
XX
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW environmental change; development; cell proliferation; differentiation;
KW elongation; survival; disease resistance; nutrient metabolism; ss.

OS Eucalyptus grandis.
XX
XX WO200042171-A1.
XX
XX 20-JUL-2000.
XX
XX 11-JAN-2000; 2000MO-US00724.
XX
XX

PR 12-JAN-1999; 99US-0228986.
PR 01-NOV-1999; 99US-0162866.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Strabala TJ, Nieuwenhuizen NJ;
XX
XX WPI; 2000-476052/41.
XX
XX
PT Isolated polynucleotide encoding a polypeptide involved in cell
PT signaling used for generating transgenic plants with modified responses
PT to external signals -
XX
PS Claim 1; Page 152-153; 527pp; English.
XX
XX
CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC tree free species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
XX
SQ Sequence 2884 BP; 856 A; 627 C; 705 G; 696 T; 0 other;

Query Match 74.8%; Score 17.2; DB 21; Length 2884;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 2 AGTGTGATGTGAGAGACCT 23
Db 2701 AGTGCCATGTGTGAGAGCCTT 2722

RESULT 15
AAAT9657
ID AAA79657 standard; cDNA; 3600 BP.

AC AAA79657;
DT 27-NOV-2000 (first entry)
DE Eucalyptus grandis cell signalling involved polynucleotide SEQ ID NO:458.
XX
XX
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW environmental change; development; cell proliferation; differentiation;
KW elongation; survival; disease resistance; nutrient metabolism; ss.

OS Eucalyptus grandis.
XX
XX WO200042171-A1.
XX
XX 20-JUL-2000.
XX
XX 11-JAN-2000; 2000MO-US00724.
XX
XX 12-JAN-1999; 99US-0228986.
XX
XX 01-NOV-1999; 99US-0162866.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Strabala TJ, Nieuwenhuizen NJ;
XX
XX

xx WPI: 2000-476052/41.
DR
xx

PT Isolated polynucleotide encoding a polypeptide involved in cell
PT signaling used for generating transgenic plants with modified responses
PT to external signals -
xx
xx

PS Claim 1; Page 203-204; 527pp; English.
xx

CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
xx

SO Sequence 3600 BP; 1026 A; 811 C; 843 G; 920 T; 0 other;

Query Match 74.8%; Score 17.2; DB 21; Length 3600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 AGTGGATGTTGAGAGAGCTT 23
||||| |||||||||
DB 3417 AGTGGCCATGTTGAGAGAGCTT 3438

Search completed: January 24, 2003, 19:45:32
Job time : 6.27201 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:32 : Search time 56.8981 Seconds
(without alignments)
11764.270 Million cell updates/sec

Title: US-09-877-935-2

Perfect score: 23

Sequence: 1 gagtggatgtgtgagagagcct 23

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hgtgo_hum:*

40: em_hgtgo_mus:*

41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 23 | 100.0 | 23 | 6 AX027825 | AX027825 Sequence |
| 2 | 23 | 100.0 | 3060 | 10 MUSVILLIN | M88454 Mus musculus |
| 3 | 23 | 100.0 | 3104 | 10 BC015267 | BC015267 Mus muscu |
| 4 | 23 | 100.0 | 192060 | 2 AC098570 | AC098570 Mus muscu |
| 5 | 21.4 | 93.0 | 1501 | 9 BC017303 | BC017303 Homo sapi |
| 6 | 21.4 | 93.0 | 2702 | 6 A07400 | A07400 H.sapiens D |
| 7 | 21.4 | 93.0 | 2702 | 6 HSYTILR | X12901 Human mRNA |
| 8 | 21.4 | 93.0 | 149109 | 2 AC101018 | AC101018 Rattus no |
| 9 | 21.4 | 93.0 | 149425 | 2 AC013320 | AC013320 Homo sapi |
| 10 | 21.4 | 93.0 | 159492 | 2 AC115173 | AC115173 Rattus no |
| 11 | 21.4 | 93.0 | 191754 | 9 AC021016 | AC021016 Homo sapi |
| 12 | 19.8 | 86.1 | 173085 | 2 AC112105 | AC112105 Rattus no |
| 13 | 18.8 | 81.7 | 18705 | 3 CEY266FA | AL023830 Caenorhab |
| 14 | 18.8 | 81.7 | 86381 | 2 AL360077 | AL360077 Homo sapi |
| 15 | 18.8 | 81.7 | 100472 | 2 AP002875 | AP002875 Homo sapi |
| 16 | 18.8 | 81.7 | 131239 | 9 HS667H12 | AL035414 Human DNA |
| 17 | 18.8 | 81.7 | 143080 | 9 AC026431 | AC026431 Homo sapi |
| 18 | 18.8 | 81.7 | 144154 | 2 AP003762 | AP003762 Oryza sat |
| 19 | 18.8 | 81.7 | 152586 | 2 AC115073 | AC115073 Mus muscu |
| 20 | 18.8 | 81.7 | 160216 | 2 AC095951 | AC095951 Rattus no |
| 21 | 18.8 | 81.7 | 166143 | 2 AL356960 | AL356960 Homo sapi |
| 22 | 18.8 | 81.7 | 167284 | 9 AC112719 | AC112719 Homo sapi |
| 23 | 18.8 | 81.7 | 169263 | 2 AC117294 | AC117294 Rattus no |
| 24 | 18.8 | 81.7 | 174255 | 2 AC129726 | AC129726 Rattus no |
| 25 | 18.8 | 81.7 | 181297 | 2 AC020675 | AC020675 Homo sapi |
| 26 | 18.8 | 81.7 | 185144 | 2 AC095764 | AC095764 Rattus no |
| 27 | 18.8 | 81.7 | 185245 | 2 AC128046 | AC128046 Rattus no |
| 28 | 18.8 | 81.7 | 185932 | 2 AP003714 | AP003714 Oryza sat |
| 29 | 18.8 | 81.7 | 186787 | 9 AC010267 | AC010267 Homo sapi |
| 30 | 18.8 | 81.7 | 187415 | 2 AC110802 | AC110802 Homo sapi |
| 31 | 18.8 | 81.7 | 229537 | 2 AC127312 | AC127312 Mus muscu |
| 32 | 18.4 | 80.0 | 4151 | 14 H2U03488 | U03488 Heliothis z |
| 33 | 18.4 | 80.0 | 4286 | 6 AR062598 | AR062598 Sequence |
| 34 | 18.4 | 80.0 | 4286 | 6 AR071501 | AR071501 Sequence |
| 35 | 18.4 | 80.0 | 4286 | 6 E10704 | E10704 DNA sequence |
| 36 | 18.4 | 80.0 | 14708 | 3 AE002952 | AE002952 Drosophila |
| 37 | 18.4 | 80.0 | 75162 | 2 AC018480 | AC018480 Drosophila |
| 38 | 18.4 | 80.0 | 112755 | 10 AC083887 | AC083887 Mus muscu |
| 39 | 18.4 | 80.0 | 138851 | 10 AL645938 | AL645938 Mouse DNA |
| 40 | 18.4 | 80.0 | 167706 | 2 AC095302 | AC095302 Rattus no |
| 41 | 18.4 | 80.0 | 169341 | 2 AC107753 | AC107753 Mus muscu |
| 42 | 18.4 | 80.0 | 209572 | 2 AC099615 | AC099615 Mus muscu |
| 43 | 18.2 | 79.1 | 1663 | 6 AX467017 | AX467017 Sequence |
| 44 | 18.2 | 79.1 | 9223 | 8 AF435120 | AF435120 Aspergill |
| 45 | 18.2 | 79.1 | 121116 | 2 AC024661 | AC024661 Homo sapi |

ALIGNMENTS

| RESULT 1 | AX027825 | 23 bp | DNA | linear | PAT 16-SEP-2000 |
|------------|---|-------------|-----|--------|-----------------|
| LOCUS | AX027825 | | | | |
| DEFINITION | Sequence 2 from Patent WO0034492. | | | | |
| ACCESSION | AX027825 | | | | |
| VERSION | AX027825.1 | GI:10188666 | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| REFERENCE | 1 (bases 1 to 23) | | | | |
| AUTHORS | Robine,S., Louvard,D., Pinto,D. and Jaissier,F. | | | | |
| TITLE | Regulatory sequences of the mouse villin gene - use in transgenesis | | | | |
| JOURNAL | Patent: WO 0034492-A 2 15-JUN-2000; | | | | |
| | ROBINE SYLVIE (FR) ; INST CURIE (FR) ; LOUWARD DANIEL (FR) ; PINTO | | | | |

FEATURES DANIEL (FR) : CENTRE NAT RECH SCIENT (FR) : JAISSER FREDERIC (FR)
 Location/Qualifiers
 source 1..23
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="oligonucleotide"
 BASE COUNT 5 a 2 c 10 g 6 t
 ORIGIN
 Query Match 100.0%: Score 23; DB 6; Length 23;
 Best Local Similarity 100.0%: Pred. No. 1.2;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GAGTGTGATGTTGAGAGAGCCT 23
 Db 1 GAGTGTGATGTTGAGAGAGCCT 23
 RESULT 2
 MUSVILLIN/c 3060 bp mRNA linear ROD 27-Apr-1993
 LOCUS MUS musculus villin protein mRNA, complete cds.
 DEFINITION M88454
 ACCESSION M88454.1 GI:202365
 VERSION villin.
 KEYWORDS Mus musculus (strain FVB) (library: lambda ZAP) adult intestinal epithelial cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 3060)
 AUTHORS Nix,D.A., Ezzeil,R.M., Leung,J., Simister,N.E., de Arruda,M.V. and Matsudaira,P.T.
 JOURNAL Unpublished (1992)
 FEATURES
 source Location/Qualifiers
 1..3060
 /organism="Mus musculus"
 /strain="FVB"
 /db_xref="taxon:10090"
 /tissue_type="intestinal epithelial"
 /dev_stage="adult"
 /tissue_1lb="lambda ZAP"
 55..2538
 /codon_start=1
 /product="villin"
 /protein_id="AAA40554.1"
 /db_xref="GI:202366"
 /translation="MTKLNQVKGSLNTTPTGQIRIEMQNVVPVPSSTFGSPFEDG
 CYVVALIHKTSSTLSYDIHWIGDSSODEGAAATYTTOMDYLLKGRAVQREYVGN
 ESETFPSYFKQGLVIRKGGVAGSMKHVETNSCDVQLLHVKGKRVNLAGVEKMSF
 NRGDVFLLDLGLKLIQWNGPESNMERLQMLAKELRDQERGRTYVGVVDEKSGD
 SPOLMAIMNHVILPRKELKAISDSVVEPAKALSLHYSDSEKGLVVEVATRLPT
 ODLKHEDCYILDDGLKIFVWKGKNAKNAOERGAALNFIKAKOYPPSOVEYON
 DGAESPITFOGLFQKRTVPNRTSLGKTHYVSAKYQKFDALTMHVDPOVAQOKM
 VDDGSEVQWRIEDELIVPESKMLGHTFGDCYLLTYLLGKQHYLLTYMQSQ
 ASODEIAASAYQAVLLDOKYNDPVOIRVTKGEPHLSIFKGRNVVVGGSKNN
 LEVPSITRLEOVGTNADNTKAEVYARATSLNSNDVFLTKTPSCCYLVGCGKSCDE
 REMAKVAVADTISRTEKOVYVEGEPANFMAALGKAPYANTKRLQDENOVITPRLEC
 SNOTGRFLATEIFEDNDLDEEDVFLVDMVOPFWIGKHAHEEKKAATTVQCYL
 KTRPGNRDLPTIIVYKQCHHEPTFTGWLAMPDFKSNSTKSYDDLKAKLNGSGMSQ
 IADENVSPKVDVFTANTSLSSGPLTFPFLQLVNKSVEDLPEGVDPSRKEEHLTDF
 TRALGMPAFAFSLPRMKQONIKKEKGLF"
 polyA_signal 3039..3045
 polyA_site 3060
 /note="putative"
 BASE COUNT 770 a 794 c 849 g 647 t
 ORIGIN
 Query Match 100.0%: Score 23; DB 10; Length 3060;
 Best Local Similarity 100.0%: Pred. No. 1.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GAGTGTGATGTTGAGAGAGCCT 23

Db 103 GAGTGTGATGTTGAGAGAGCCT 81
 RESULT 3
 BC015267/c 3104 bp mRNA linear ROD 07-AUG-2002
 LOCUS MUS musculus, villin, clone MGC:18506 IMAGE:4236751, mRNA, complete
 DEFINITION cds.
 ACCESSION BC015267
 VERSION BC015267.1 GI:15929677
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 3104)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdepall.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 25 Row: m Column: 12
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6678572.
 FEATURES
 source Location/Qualifiers
 1..3104
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:18506 IMAGE:4236751"
 /tissue_type="Kidney, normal. 5 month old male mouse."
 /clone_1lb="NCI_CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 41..2524
 /codon_start=1
 /product="villin"
 /protein_id="AAH15267.1"
 /db_xref="GI:15929678"
 /db_xref="LocusID:22349"
 /translation="MTKLNQVKGSLNTTPTGQIRIEMQNVVPVPSSTFGSPFEDG
 CYVVALIHKTSSTLSYDIHWIGDSSODEGAAATYTTOMDYLLKGRAVQREYVGN
 ESETFPSYFKQGLVIRKGGVAGSMKHVETNSCDVQLLHVKGKRVNLAGVEKMSF
 NRGDVFLLDLGLKLIQWNGPESNMERLQMLAKELRDQERGRTYVGVVDEKSGD
 SPOLMAIMNHVILPRKELKAISDSVVEPAKALSLHYSDSEKGLVVEVATRLPT
 ODLKHEDCYILDDGLKIFVWKGKNAKNAOERGAALNFIKAKOYPPSOVEYON
 DGAESPITFOGLFQKRTVPNRTSLGKTHYVSAKYQKFDALTMHVDPOVAQOKM
 VDDGSEVQWRIEDELIVPESKMLGHTFGDCYLLTYLLGKQHYLLTYMQSQ
 ASODEIAASAYQAVLLDOKYNDPVOIRVTKGEPHLSIFKGRNVVVGGSKNN
 LEVPSITRLEOVGTNADNTKAEVYARATSLNSNDVFLTKTPSCCYLVGCGKSCDE
 REMAKVAVADTISRTEKOVYVEGEPANFMAALGKAPYANTKRLQDENOVITPRLEC
 SNOTGRFLATEIFEDNDLDEEDVFLVDMVOPFWIGKHAHEEKKAATTVQCYL
 KTRPGNRDLPTIIVYKQCHHEPTFTGWLAMPDFKSNSTKSYDDLKAKLNGSGMSQ
 IADENVSPKVDVFTANTSLSSGPLTFPFLQLVNKSVEDLPEGVDPSRKEEHLTDF
 TRALGMPAFAFSLPRMKQONIKKEKGLF"
 BASE COUNT 803 a 795 c 851 g 655 t

ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 3104;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGACCT 23
 ||||||||||||||||||
 DB 89 GAGTGTGATGTTGAGAGACCT 67

RESULT 4
 AC098570/c
 LOCUS
 DEFINITION Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered pieces.
 AC098570
 VERSION AC098570.2 GI:22325297
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 192060)
 Birren,B., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-278N11
 JOURNAL
 TITLE Unpublished
 AUTHORS 2 (bases 1 to 192060)
 REFERENCES
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferrelia,P., Fitzhugh,W., Gage,D., Galagan,J., Gaidyne,S., Glndse,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nobu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 192060)
 Birren,B., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferrelia,P., Fitzgerald,M., Gage,D., Galagan,J., Gaidyne,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnpbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission

JOURNAL

COMMENT

Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 20, 2002 this sequence version replaced gi:16356898.
 All repeats were identified using RepeatMasker:
 Smit,A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/MW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L14796
 Center clone name: 278.N.11
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 190464 bases at least Q40
 Consensus quality: 191061 bases at least Q30
 Consensus quality: 191345 bases at least Q20
 Insert size: 163000; agarose-1p
 Insert size: 191560; sum-of-ctrls
 Quality coverage: 12.4 in Q20 bases; agarose-1p
 Quality coverage: 10.5 in Q20 bases; sum-of-ctrls

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 4921: contig of 4921 bp in length
 * 4922 5021: gap of 100 bp
 * 5022 15200: contig of 10179 bp in length
 * 15201 15300: gap of 100 bp
 * 15301 78692: contig of 63392 bp in length
 * 78693 78792: gap of 100 bp
 * 78793 116118: contig of 37326 bp in length
 * 116119 116218: gap of 100 bp
 * 116219 174491: contig of 58273 bp in length
 * 174492 174591: gap of 100 bp
 * 174592 192060: contig of 17469 bp in length.
 Location/Qualifiers
 1..192060
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RP23-278N11"
 /clone_lib="RP23-278N11"
 /clone_lib="RP23-278N11"
 1..4921
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 5022..15200
 /note="assembly-fragment"
 15301..78692
 /note="assembly-fragment"
 78793..116118
 /note="assembly-fragment"
 116219..174491
 /note="assembly-fragment"
 174592..192060
 /note="assembly-fragment
 clone_end:T7
 vector_side:right"
 BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others
 ORIGIN

Query Match 100.0%; Score 23; DB 2; Length 192060;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23
|||||
Db 49339 GAGTGTGATGTTGAGAGAGCCT 49317

RESULT 5
BC017303 1501 bp mRNA linear PRI 09-NOV-2001
LOCUS Homo sapiens, similar to villin 1, clone MGC:29534 IMAGE:5087169,
DEFINITION mRNA, complete cds.
ACCESSION BC017303
VERSION BC017303.1 GI:16878196
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1501)
AUTHORS Strausberg, R.
TITLE Direct Submissions
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcaps-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland:
Web site: http://www.nisc.nih.gov/nisc_mgc.html
Contact: nisc_mgc@nsl.nsl.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stanlipop, S., Thomas, J.J.,
Tongson, E.E., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, M.A.,
Zhang, L., H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 39 Row: 0 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6005943.
location/Qualifiers
1. 1501
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:29534 IMAGE:5087169"
/issue_type="Liver, hepatocellular carcinoma"
/clone_1db="NIH_MGC_100"
/lab_host="DH10B"
/note="Vector: pOTB7"
30. 1295
/codon_start=-1
/product="Similar to villin 1"
/protein_id="AAH17303.1"
/db_xref="GI:16878197"
/translation="MTKLSAIVGSGSLNITPGIOWIRIAMQVVPVPSSTFGSFPDQ
CYIIAIIHKTASSLSDIHYIKGODSSLDGGAAATYTTOMDFLKRAYOHREVOGN
ESBAFRTGYOGVILIRKGVASGKSHETNSYOVRLHAKGRNVAVAGEVMSKSF
NRGDFVLDLGIKLIQWNGPESTRMEKRLMTLAKELRDERGGRITVGVVDENELA
SPKLEVMNHLKRRRLKAAPDVTVEPALKALAHVSDSEGNLVVEVATRPIT
QDLSHEDCVIIDGGGLKIVYVKGKANEKAGSHALNFIKAOYPSSTOVEON
DGAESAVFOQLFOKWTASNTSGIKTHHTVGSVCEGAGAVREPGRSMARRATWSTT
HPSLVTCTFEDFTAGSGVLADGDVDTKL"

BASE COUNT 394 a 378 c 468 g 261 t

Query Match 93.0%; Score 21.4; DB 9; Length 1501;
Best Local Similarity 95.7%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23
|||||
Db 78 GAGTGTGATGTTGAGAGAGCCT 56

RESULT 6
A07400/c 2702 bp DNA linear PAT 20-JUL-1993
LOCUS H.sapiens DNA for villin.
DEFINITION A07400
ACCESSION A07400.1 GI:412171
VERSION villin.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2702)
AUTHORS Arpin, M., Pringault, E., Garcia, A. and Louvard, D.
TITLE Specific nucleic-acid fragments of the human villin gene and their
application to diagnostic ends
JOURNAL PASTEUR
INSTITUT PASTEUR
location/Qualifiers
1. 2702
/organism="Homo sapiens"
/db_xref="taxon:9606"
25. 2508
/codon_start=-1
/product="villin"
/protein_id="CAA00664.1"
/db_xref="GI:412172"
/db_xref="SWISS-PROT:P09327"
/translation="MTKLSAIVGSGSLNITPGIOWIRIAMQVVPVPSSTFGSFPDQ
CYIIAIIHKTASSLSDIHYIKGODSSLDGGAAATYTTOMDFLKRAYOHREVOGN
ESBAFRTGYOGVILIRKGVASGKSHETNSYOVRLHAKGRNVAVAGEVMSKSF
NRGDFVLDLGIKLIQWNGPESTRMEKRLMTLAKELRDERGGRITVGVVDENELA
SPKLEVMNHLKRRRLKAAPDVTVEPALKALAHVSDSEGNLVVEVATRPIT
QDLSHEDCVIIDGGGLKIVYVKGKANEKAGSHALNFIKAOYPSSTOVEON
DGAESAVFOQLFOKWTASNTSGIKTHHTVGSVCEGAGAVREPGRSMARRATWSTT
HPSLVTCTFEDFTAGSGVLADGDVDTKL"

BASE COUNT 681 a 716 c 780 g 524 t 1 others

ORIGIN

Query Match 93.0%; Score 21.4; DB 6; Length 2702;
Best Local Similarity 95.7%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTTGAGAGAGCCT 23
|||||
Db 73 GAGTGTGATGTTGAGAGAGCCT 51

RESULT 7
HSVILR/c 2702 bp mRNA linear PRI 12-SEP-1993
LOCUS Human mRNA for villin.
DEFINITION X12901
ACCESSION X12901.1 GI:37842
KEYWORDS actin-binding protein; calcium binding protein; villin.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 149109)
 Bliren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouknight,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cook,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gadyana,S., Gird,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,V., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Notman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V., Roman,J., Rossetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Strauss,N., Subramaniam,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and zody,M.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 COMMENT

Direct Submission
 Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17059792.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMask

```

* 70110 77196: contig of 7087 bp in length
* 77197 77296: gap of 100 bp
* 77297 99453: contig of 22157 bp in length
* 99454 99553: gap of 100 bp
* 99554 102879: contig of 3326 bp in length
* 102880 102879: gap of 100 bp
* 102980 120352: contig of 17373 bp in length
* 120353 120452: gap of 100 bp
* 120453 131546: contig of 11094 bp in length
* 131547 131646: gap of 100 bp
* 131647 149109: contig of 17463 bp in length.
Location/Qualifiers
1. 149109
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_1lb="RP32-328P7"
/clone_1lb="RP32-328P7"
BASE COUNT 39036 a 34316 c 35185 g 39719 t 853 others
ORIGIN
Query Match 93.0% Score 21.4; DB 2; Length 149109;
Best Local Similarity 95.7% Pred: No. 8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGTGATGTGAGAGACCT 23
|||||
Db 105687 GAGTGTGATGTGAGAGACCT 105665

RESULT 9
AC013320/c 149425 bp DNA linear HTG 18-NOV-2000
LOCUS Homo sapiens chromosome 2 clone RP11-36003 map 2, WORKING DRAFT
DEFINITION
AC013320 149425 bp DNA linear HTG 18-NOV-2000
AC013320 Homo sapiens chromosome 2 clone RP11-36003 map 2, WORKING DRAFT
SEQUENCE 19 unordered pieces.
AC013320
AC013320.7 GI:11225433
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149425)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-36003
Unpublished
2 (bases 1 to 149425)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckert,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,M., Forrest,C., Funke,R., Gage,D.,
Galsburg,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2000 this sequence version replaced gi:10864221.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L3776
Center clone name: 360_O_3
----- Summary Statistics
Sequencing vector: M13; M77815; 12% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 140887 bases at least Q40
Consensus quality: 144879 bases at least Q30
Consensus quality: 146417 bases at least Q20
Insert size: 228000; agarose-fp
Insert size: 147625; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 ba.
NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
8145 8244: gap of 100 bp
8245 9350: contig of 1106 bp in length
9351 9450: gap of 100 bp
9451 11335: contig of 1885 bp in length
1136 11435: gap of 100 bp
1136 14325: contig of 2890 bp in length
1436 14425: gap of 100 bp
14426 17305: contig of 2880 bp in length
17306 17405: gap of 100 bp
17406 20166: contig of 2761 bp in length
20167 20266: gap of 100 bp
20267 23474: contig of 3208 bp in length
23475 23574: gap of 100 bp
23575 26403: contig of 2829 bp in length
26404 26503: gap of 100 bp
26504 29899: contig of 3396 bp in length
29900 29999: gap of 100 bp
30000 34929: contig of 4930 bp in length
34930 35029: gap of 100 bp
35030 41463: contig of 6434 bp in length
41464 41563: gap of 100 bp
41564 47890: contig of 6327 bp in length
47891 47990: gap of 100 bp
47991 56092: contig of 8102 bp in length
56093 56192: gap of 100 bp
56193 63063: contig of 6871 bp in length
63064 63163: gap of 100 bp
63164 70819: contig of 7656 bp in length
70820 70919: gap of 100 bp
70920 82482: contig of 11563 bp in length
82483 82582: gap of 100 bp
82583 98645: contig of 16063 bp in length
98646 98745: gap of 100 bp
98746 117782: contig of 19037 bp in length
117783 117882: gap of 100 bp
117883 149425: contig of 31543 bp in length.
Location/Qualifiers
1. 149425
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-36003"
/clone_1lb="RP32-328P7"
/clone_1lb="RP32-328P7"
1. 8144
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
8245. 9350
/note="assembly_fragment"

```

misc_feature

misc_feature


```

* 2396 2495: gap of unknown length
* 2496 3600: contig of 1105 bp in length
* 3601 3700: gap of unknown length
* 3701 5021: contig of 1321 bp in length
* 5022 5121: gap of unknown length
* 5122 6727: contig of 1606 bp in length
* 6728 6827: gap of unknown length
* 6828 8608: contig of 1781 bp in length
* 8609 10003: contig of 1295 bp in length
* 10004 10103: gap of unknown length
* 10104 11211: contig of 1108 bp in length
* 11212 11311: gap of unknown length
* 11312 12518: contig of 1207 bp in length
* 12519 12618: gap of unknown length
* 12619 13843: contig of 1225 bp in length
* 13844 13943: gap of unknown length
* 13944 15404: contig of 1461 bp in length
* 15405 15505: gap of unknown length
* 15506 17647: contig of 2143 bp in length
* 17648 17747: gap of unknown length
* 17748 19706: contig of 1959 bp in length
* 19707 19806: gap of unknown length
* 19807 21576: contig of 1770 bp in length
* 21577 21676: gap of unknown length
* 21677 22906: contig of 1230 bp in length
* 22907 23006: gap of unknown length
* 23007 25086: contig of 2080 bp in length
* 25087 25186: gap of unknown length
* 25187 26832: contig of 1646 bp in length
* 26833 26932: gap of unknown length
* 26933 28710: contig of 1778 bp in length
* 28711 28810: gap of unknown length
* 28811 30606: contig of 1796 bp in length
* 30607 30706: gap of unknown length
* 30707 33113: contig of 2407 bp in length
* 33114 33213: gap of unknown length
* 33214 35361: contig of 2148 bp in length
* 35362 37520: contig of 2059 bp in length
* 37521 37620: gap of unknown length
* 37621 40471: contig of 2851 bp in length
* 40472 40571: gap of unknown length
* 40572 42584: contig of 2013 bp in length
* 42585 42684: gap of unknown length
* 42685 45547: contig of 2863 bp in length
* 45548 45647: gap of unknown length
* 45648 47838: contig of 2191 bp in length
* 47839 47938: gap of unknown length
* 47939 51229: contig of 3291 bp in length
* 51230 51329: gap of unknown length
* 51330 54005: contig of 2676 bp in length
* 54006 54105: gap of unknown length
* 54106 56683: contig of 2578 bp in length
* 56684 56783: gap of unknown length
* 56784 59361: contig of 2578 bp in length
* 59362 59461: gap of unknown length
* 59462 61588: contig of 2127 bp in length
* 61589 61688: gap of unknown length
* 61689 65712: contig of 4024 bp in length
* 65713 65812: gap of unknown length
* 65813 69339: contig of 3527 bp in length
* 69340 69439: gap of unknown length
* 69440 72570: contig of 3131 bp in length
* 72571 72670: gap of unknown length
* 72671 76896: contig of 4228 bp in length
* 76897 76996: gap of unknown length
* 76997 80760: contig of 3762 bp in length
* 80761 80860: gap of unknown length
* 80861 86313: contig of 5453 bp in length
* 86314 86413: gap of unknown length
* 86414 92231: contig of 5818 bp in length
* 92232 92331: gap of unknown length

```

```

* 92332 99692: contig of 7361 bp in length
* 99693 99792: gap of unknown length
* 99793 106178: contig of 6386 bp in length
* 106179 106278: gap of unknown length
* 106279 114304: contig of 8026 bp in length
* 114305 114404: gap of unknown length
* 114405 122112: contig of 7708 bp in length
* 122113 122212: gap of unknown length
* 122213 129642: contig of 7430 bp in length
* 129643 129742: gap of unknown length
* 129743 138684: contig of 8942 bp in length
* 138685 138784: gap of unknown length
* 138785 149662: contig of 10878 bp in length
* 149663 149762: gap of unknown length
* 149763 159492: contig of 9730 bp in length.
* 159492 location/Qualifiers
  1.159492
  /organism="Rattus norvegicus"
  /db_xref="taxon:10116"
  /clone="CH230-286017"

BASE COUNT 39083 a 36447 c 36055 g 40464 t 7443 others
ORIGIN
Query Match 93.0%; Score 21.4; DB 2; Length 159492;
Best Local Similarity 95.7%; Pred. No. 8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGTGGTGGTGGAGAGCCT 23
Db 13423 GAGTGGTGGTGGAGAGCCT 13445

RESULT 11
AC021016/c 191754 bp DNA linear PRI 09-MAY-2001
LOCUS AC021016
DEFINITION Homo sapiens BAC clone RP11-378A13 from 2, complete sequence.
ACCESSION AC021016
VERSION AC021016.4 GI:11120952
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 191754)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 191754)
AUTHORS Cordes,M., Maupin,R., Hawkins,M. and Boyer,E.
TITLE The sequence of Homo sapiens BAC clone RP11-378A13
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 191754)
AUTHORS Waterston,R.H.
TITLE Direct Submmission
JOURNAL Submitted (12-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 191754)
AUTHORS Waterston,R.
TITLE Direct Submmission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington
University 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 191754)
AUTHORS Waterston,R.
TITLE Direct Submmission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gl:7630969.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center

```


Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0378A13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Teleno, M., Cataneese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-146N10; the clone sequenced to the right is RP11-36003. Actual start of this clone is at base position 1 of RP11-378A13; actual end is at base position 191754 of RP11-378A13.

The sequence RP11-378A13 from 156426 to 156433 is derived from a single m13 subclone. Size of the region was confirmed by PCR from BAC DNA.

FEATURES

SOURCE

1. 191754 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-378A13"
/clone_id="RPCI-11"
misc_feature 1. 224 /note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"
misc_feature 1. 224 /note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"
misc_feature 1. 224 /note="similar to EST A1457236 (NID:g4310105) tk01d06.x1"
misc_feature 1. 216 /note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"
misc_feature 1. 216 /note="similar to EST AW020425 (NID:g5673955) df09e11.y1"
misc_feature 2. 223 /note="similar to EST AA175052 (NID:g1756173) ms82f02.r1"
misc_feature 2. 223 /note="similar to EST AV079022 (NID:g5198850)"
misc_feature 2. 223 /note="similar to EST AW538472 (NID:g7180889)"
misc_feature 2. 49 /note="similar to EST AV330904 (NID:g6370956)"
misc_feature 7. 224 /note="similar to EST A1454294 (NID:g4292719)"
misc_feature 7. 224

misc_feature 12. 223 /note="similar to EST AW433991 (NID:g6365298)"
misc_feature 20. 223 /note="similar to EST AW059787 (NID:g6652109)"
misc_feature 60. 220 /note="similar to EST A1709077 (NID:g4998853) as86d11.x1"
misc_feature 67. 223 /note="similar to EST AA870364 (NID:g2655809) vq44e05.r1"
repeat_region 447. 660 /note="similar to EST AV330904 (NID:g6370956)"
repeat_region 1182. 1425 /rpt_family="Alu"
repeat_region 1442. 1583 /rpt_family="Alu"
repeat_region 1820. 1928 /rpt_family="MERL_type"
misc_feature 1948. 2347 /rpt_family="Alu"
repeat_region 2162. 2293 /note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"
repeat_region 2297. 2587 /rpt_family="Alu"
repeat_region 2597. 2804 /rpt_family="Alu"
misc_feature 2606. 2923 /rpt_family="Alu"
misc_feature 2630. 3135 /note="similar to EST AW351965 (NID:g6849698)"
misc_feature 2804. 3242 /note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"
misc_feature 2976. 3245 /note="similar to EST A1916611 (NID:g5536466) wa28f08.x1"
repeat_region 3097. 3284 /note="similar to EST A1401669 (NID:g4244756) th24e10.x1"
misc_feature 3546. 3781 /rpt_family="MIR"
repeat_region 3848. 4127 /note="similar to EST AA102341 (NID:g1647297) zm26h05.r1"
repeat_region 4187. 4469 /rpt_family="Alu"
repeat_region 4474. 4690 /rpt_family="Alu"
misc_feature 4723. 5241 /rpt_family="Alu"
misc_feature 4931. 5413 /note="similar to EST AW444974 (NID:g6386736)"
repeat_region 6006. 6137 /note="similar to EST AA100813 (NID:g1647248) zm26h05.s1"
repeat_region 6243. 6394 /rpt_family="MIR"
misc_feature 6967. 7092 /rpt_family="MIR"
misc_feature 7935. 8054 /note="similar to EST AW461516 (NID:g7031684)"
repeat_region 8294. 8433 /note="similar to EST AW461516 (NID:g7031684)"
misc_feature 8533. 8604 /rpt_family="MERL_type"
misc_feature 8556. 8749 /note="similar to EST A1391439 (NID:g4217443) tf96e02.x1"
misc_feature 9362. 9733 /note="similar to EST A1391439 (NID:g4217443) tf96e02.x1"
misc_feature 9476. 9723 /note="similar to EST A1391439 (NID:g4217443) tf96e02.x1"
misc_feature 9997. 10265 /note="similar to EST AA595679 (NID:g2411029) nk66f01.s1"
misc_feature 10003. 10482 /note="similar to EST A1634626 (NID:g4685956) tz30g01.x1"
misc_feature 10003. 10482 /note="similar to EST A1828100 (NID:g5448771) wk31g06.x1"
misc_feature 10003. 10482 /note="similar to EST A1890574 (NID:g5595738) wm87e04.x1"
misc_feature /note="similar to EST A1978960 (NID:g5803990) wr8e06.x1"

```

msc_feature 10003..10482
             /note="similar to EST AL047914 (NID:g4728747)"
msc_feature 10003..10482
             /note="similar to EST AM027349 (NID:g5886105) w73e09.x1"
msc_feature 10003..10457
             /note="similar to EST A1972200 (NID:g5769026) w763n01.x1"
msc_feature 10003..10421
             /note="similar to EST W48859 (NID:g1336988) zc42c08.s1"
msc_feature 10003..10330
             /note="similar to EST AM074215 (NID:g6029213) xb09612.x1"
msc_feature 10003..10249
             /note="similar to EST AA143563 (NID:g1712935) zc32d10.s1"
msc_feature 10003..10247
             /note="similar to EST A1583619 (NID:g4569516) ts16d09.x1"
msc_feature 10003..10240
             /note="similar to EST A1583619 (NID:g4569516) ts16d09.x1"

Query Match 93.0%; Score 21.4; DB 9; Length 191754;
Best Local Similarity 95.7%; Pred. No. 8;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAGTGTGATGTTGAGAGACCT 23
Db 169701 GGGTGTGATGTTGAGAGACCT 169679

RESULT 12
AC112105
LOCUS
DEFINITION
AC112105 173085 bp DNA linear HTG 17-JUL-2002
Rattus norvegicus clone CH230-301N9, *** SEQUENCING IN PROGRESS
***, 58 unordered pieces.
AC112105
VERSION
AC112105.3 GI:21744618
HTG: HTGS PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 173085)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alshrocks,S.L., Anaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,J., Blinze,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowe,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Burch,C., Burch,P., Burke,C., Butrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Jolyet,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kralevic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Lousaged,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Maxwell,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Nickerson,E., Nwokenko,S., Ogih,M., Okunou,G.,
Oradunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pichens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,I., Rolle,M., Ruiz,S., Savely,G.,
Scheer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,

```

```

Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 173085)
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 173085)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:20303309.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSIW
Center clone name: CH230-301N9
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 123259 bases at least Q40
Consensus quality: 129338 bases at least Q30
Consensus quality: 134476 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1043 1042: contig of 1042 bp in length
1143 1142: gap of unknown length
1143 2142: contig of 1000 bp in length
2143 2242: gap of unknown length
2243 3440: contig of 1198 bp in length
3441 3540: gap of unknown length
3541 4828: contig of 1288 bp in length
4829 4929: gap of unknown length
4929 6007: contig of 1079 bp in length
6008 6108: gap of unknown length
6108 7456: contig of 1349 bp in length
7457 7556: gap of unknown length
7557 8803: contig of 1247 bp in length
8804 10250: contig of 1347 bp in length
10251 10350: gap of unknown length
10351 11457: contig of 1107 bp in length
11458 11557: gap of unknown length
11558 12846: contig of 1289 bp in length
12847 14228: gap of unknown length
14228 14329: contig of 1282 bp in length
14329 14329: gap of unknown length
14329 15874: contig of 1546 bp in length
15875 15974: gap of unknown length
15975 17615: contig of 1641 bp in length
17616 17715: gap of unknown length

```

```

* 17716 19098: contig of 1383 bp in length
* 19099 19198: gap of unknown length
* 19199 21275: contig of 2077 bp in length
* 21276 21375: gap of unknown length
* 21376 22377: contig of 1002 bp in length
* 22378 22477: gap of unknown length
* 22478 23752: contig of 1275 bp in length
* 23753 23852: gap of unknown length
* 23853 25173: contig of 1321 bp in length
* 25174 25273: gap of unknown length
* 25274 26931: contig of 1658 bp in length
* 26932 27031: gap of unknown length
* 27032 28858: contig of 1827 bp in length
* 28859 30192: gap of unknown length
* 28959 30292: contig of 1234 bp in length
* 30193 32429: gap of unknown length
* 30293 32429: contig of 2137 bp in length
* 32430 32529: gap of unknown length
* 32530 34037: contig of 1508 bp in length
* 34038 34137: gap of unknown length
* 34138 36540: contig of 2403 bp in length
* 36541 36641: gap of unknown length
* 36641 38123: contig of 1483 bp in length
* 38124 38224: gap of unknown length
* 38224 39986: contig of 1763 bp in length
* 39987 40086: gap of unknown length
* 40087 41897: contig of 1811 bp in length
* 41898 41997: gap of unknown length
* 41998 44140: contig of 2143 bp in length
* 44141 44240: gap of unknown length
* 44241 45410: contig of 1170 bp in length
* 45411 45510: gap of unknown length
* 45511 47629: contig of 2119 bp in length
* 47630 47729: gap of unknown length
* 47730 50367: contig of 2638 bp in length
* 50368 50467: gap of unknown length
* 50468 52623: contig of 2156 bp in length
* 52624 52723: gap of unknown length
* 52724 54648: contig of 1925 bp in length
* 54649 54748: gap of unknown length
* 54749 57379: contig of 2631 bp in length
* 57380 57479: gap of unknown length
* 57480 60063: contig of 2584 bp in length
* 60064 60163: gap of unknown length
* 60164 62589: contig of 2426 bp in length
* 62590 62689: gap of unknown length
* 62690 64485: contig of 1796 bp in length
* 64486 64585: gap of unknown length
* 64586 66646: contig of 2061 bp in length
* 66647 66746: gap of unknown length
* 66747 69760: contig of 3014 bp in length
* 69761 69860: gap of unknown length
* 69861 73271: contig of 3411 bp in length
* 73272 73372: gap of unknown length
* 73373 78287: contig of 4916 bp in length
* 78288 78387: gap of unknown length
* 78388 82141: contig of 3754 bp in length
* 82142 82241: gap of unknown length
* 82242 86068: contig of 3827 bp in length
* 86069 86168: gap of unknown length
* 86169 90383: contig of 4215 bp in length
* 90384 90483: gap of unknown length
* 90484 94073: contig of 3590 bp in length
* 94074 94173: gap of unknown length
* 94174 98119: contig of 3946 bp in length
* 98120 98219: gap of unknown length
* 98220 102403: contig of 4184 bp in length
* 102404 102503: gap of unknown length
* 102504 106962: contig of 4459 bp in length
* 106963 107062: gap of unknown length
* 107063 111515: contig of 4453 bp in length
* 111516 115636: contig of 4021 bp in length

```

```

* 115637 115736: gap of unknown length
* 115737 119071: contig of 3335 bp in length
* 119072 119171: gap of unknown length
* 119172 123459: contig of 4288 bp in length
* 123460 123559: gap of unknown length
* 123560 129565: contig of 6006 bp in length
* 129566 129665: gap of unknown length
* 129666 135079: contig of 5414 bp in length

Query Match      86.1%; Score 19.8; DB 2; Length 173085;
Best Local Similarity 91.3%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 GAGTGTGATGTTGAGAGCCT 23
Db 102276 GAGTGTGTTGTGAGAGTCT 102298

RESULT 13
CEY26E6A/c
LOCUS      CEY26E6A      18705 bp      DNA      linear      INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid Y26E6A, complete sequence.
ACCESSION  AL023830 AL008874
VERSION     AL023830.1 GI:3217845
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.

REFERENCE   1
            none.
            Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
            Science 282 (5396), 2012-2018 (1998)
            MEDLINE 99069613
            PUBMED 9851916
            REMARK The C. elegans Sequencing Consortium.
            REFERENCE 2 (bases 1 to 18705)
            AUTHORS Lennard,N.
            TITLE Direct Submission
            JOURNAL Submitted (09-JUN-1998) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jesus@anger.ac.uk or rwh@nematode.wustl.edu
            On May 14, 2001 this sequence version replaced gi:2612776.
            Coding sequences below are predicted from computer analysis, using
            predictions from GeneFINDER (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone Y26E6A.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone K02B9 is at 18602 in this sequence. The
            true right end of clone R21H8 is at 104 in this sequence. The start
            of this sequence (1..104) overlaps with the end of sequence Z78546.
            The end of this sequence (18602..18705) overlaps with the start of
            sequence Z69663.
            For a graphical representation of this sequence and its analysis
            see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=Y26E6A
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            Location/Qualifiers
                1..18705
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"

```

FEATURES

source

```
/chromosome="X"
/clone="Y26E6A"
join(complement(1130..1285),complement(910..1064),
complement(737..863),complement(278546..1:37732..37919),
complement(278546..1:37485..37686),
complement(278546..1:37082..37381),
complement(278546..1:37082..37381))
/gene="T21H8.4"
join(complement(1130..1285),complement(910..1064),
complement(737..863),complement(278546..1:37732..37919),
complement(278546..1:37485..37686),
complement(278546..1:37082..37381))
/gene="T21H8.4"
/note="similar to 7Tm receptor like"
/codon_start=1
/protein_id="CA19474.1"
/db_xref="GI:3880747"
/db_xref="SPPREMBL:062420"
/translation="MECETRGOLYHPAYLAIIFGSGYITPILITLKIVAFK
VYHPNIVFLMLNVSCLLGLTAMSAIPLMIRPDLKTYFCISKIGT
FLFAPCLVTTSHAGILERSMATYSVKNEROGALITLAVAVAAATLFIILP
EDDEMITTCLTFPSASKISGRVYMEFOLLDAVISYHLVYRNKNDKSGTS
LSROPORNEWVKTLKQVTPLLISNTGIVYITVSPFLRYKDYLPNNYEITANLF
IMHPMFMTSLIVELMIGKKRQDIHMDMASODPDDOPIHMEMNDVHFGSRL
KAQARANKRPLSANMTSITLTKFKKQTTVAASIPSPSS"
complement(join(9082..9405,9771..10229,10368..10826))
/gene="Y26E6A.2"
complement(join(9082..9405,9771..10229,10368..10826))
/gene="Y26E6A.2"
/codon_start=1
/protein_id="CA19473.1"
/db_xref="GI:3880746"
/db_xref="SPPREMBL:062419"
/translation="MADLDELFRVAFPEPPDOAKDRVNNEMKEDMELLTWTVE
EISINGPTPRYRCLIEOLOGETFEKSSALRGVGFETLTDIHLCDTMAHLT
KMVPSHLISRDNEFRPKSKNVEFKIIBDDYKLPDLKIPRMITQIMORSYKRAQ
SLAITEAANNITRONSLSKVPDLVILDFSSGEMNVKRLKNOQSDSDSKC
LAINSEVENKKTQDLNRTTILKIVTLDPEPECEAVHNHVGSRFKRIIVKQFTE
NLKMWMEPEKEVEYEVNALTDFIPFDIKTISIRDLNRDGHGRKSAKHFVIN
CALNIDTFERKALRDMDNGHREVGWVILKQFIPTPGRRHRPRTSPRKGLERTS
ELNSOMPGEITEYEVHETN"
complement(join(13206..13571,13624..13776,13855..14061,
14129..14229,14280..14528,14753..15026))
/gene="Y26E6A.1"
complement(join(13206..13571,13624..13776,13855..14061,
14129..14229,14280..14528,14753..15026))
/gene="Y26E6A.1"
/note="cDNA EST yk229h9.3 comes from this gene"
cDNA EST yk458i3.3 comes from this gene
cDNA EST yk497g4.3 comes from this gene
cDNA EST yk229h9.5 comes from this gene
cDNA EST yk458i3.5 comes from this gene
cDNA EST yk497g4.5 comes from this gene
cDNA EST yk516i7.5 comes from this gene
cDNA EST yk567h10.5 comes from this gene"
/codon_start=1
/protein_id="CA19472.1"
/db_xref="GI:3880745"
/db_xref="SPPREMBL:062418"
/translation="MLPVANNPEQANBEELRQADIDPOARNHNRNRRRREME
RANRNLQADRAQARERGPAGRAQGRARAREALRANNAQALQPDAAVLEA
LDEINRIDEDEQVAPNRADVPQVLEFANLPRILALPDHQAQANONLAMAHLL
APENYIRVARREQROSGFCIGKSSVHNAFLNHDTCASKACGACIMDMENHFLS
STYSLRQOLYKLEIRELEFEFARVNDPERLEHMEGACITCAHKLACMPNORVY
OIHANKIEKTVAHPEOFLELETKEDGILETSKTAHPDSIOCCIDGDPDETQOR
HGTWYERKDYVEEQLEKRYLEETIEKILVYRSHREVTLLILPMDKLACIEE
LKNSPRIARIQIRVHCIDPIRGVPLEGEAEIEDDEDINAAYQAAEEFQLGIT"
```

BASE COUNT
ORIGIN

6444 a 2981 c 3321 g 5959 t

Query Match 81.7%; Score 18.8; DB 3; Length 18705;
Best Local Similarity 90.98; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

FEATURES
source

```
QY 1 GAGTGTGATGTTGAGAGACC 22
II |IIIIIIIIIIIIIIIIII
Db 7751 GATTGGTGATGTGAGAGCC 7730

RESULT 14
AL360077/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
-----
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi:9801102.
-----
Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
Project Information
Center project name: bk501D19
-----
Summary Statistics
Sequencing program: XGAP4; version 4.5
Assembly vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 81882 bases at least Q40
Consensus quality: 83656 bases at least Q30
Consensus quality: 84751 bases at least Q20
Insert size: 85681; sum-of-contigs
Insert size: 102977; 7.3% error; agarose-fp
Quality coverage: 4.03x in Q20 bases; sum-of-contigs Quality
Coverage: 3.60x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 12686: contig of 12686 bp in length
12687 12786: gap of 100 bp
12787 16897: contig of 4111 bp in length
16898 16997: gap of 100 bp
16998 30118: contig of 13121 bp in length
30119 30218: gap of 100 bp
30219 33053: contig of 2835 bp in length
33054 33153: gap of 100 bp
33154 52695: contig of 19542 bp in length
52696 52795: gap of 100 bp
52796 59088: contig of 6293 bp in length
59089 59188: gap of 100 bp
59189 73338: contig of 14150 bp in length
73339 73438: gap of 100 bp
73439 86381: contig of 12943 bp in length.
Location/Qualifiers
1. 86381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="CTA-501D19"
/clone_lib="CIT978SK-A2"
```

```

misc_feature      1..12686
                  /note="assembly_fragment:00068"
misc_feature      12787..16897
                  /note="assembly_fragment:00338"
misc_feature      16998..30118
                  /note="assembly_fragment:00394"
misc_feature      30219..33053
                  /note="assembly_fragment:00465"
misc_feature      33154..52695
                  /note="assembly_fragment:00486"
misc_feature      52796..59088
                  /note="assembly_fragment:00701"
misc_feature      59189..73338
                  /note="assembly_fragment:00876"
misc_feature      73439..86381
                  /note="assembly_fragment:00891"
BASE COUNT      24179 a 19215 c 17718 g 24568 t 701 others
ORIGIN

```

```

Query Match      81.7%: Score 18.8; DB 2; length 86381;
Best Local Similarity 90.9%: Pred. No.1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 AGTGGTGAATGTTGAGAGACCT 23
        |||||
Db      45356 AGTGGTGAATGTTGAGAGACCT 45335

```

```

RESULT 15
AP002875      100472 bp DNA linear HTG 07-OCT-2000
LOCUS      Homo sapiens chromosome 4 clone 390M07 map 4q16-q20, *** SEQUENCING
DEFINITION      IN PROGRESS ***, 2 ordered pieces.
ACCESSION      AP002875
VERSION      AP002875.1 GI:10716811
KEYWORDS      HTG: HTGS_PHASE2.
SOURCE      Homo sapiens DNA, clone:390M07.
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 100472)
JOURNAL      Tsai,S.F.

```

```

COMMENT      Direct Submission
              Submitted (06-OCT-2000) Shih-Peng Tsai, National Yang-Ming
              University, Institute of Genetics, 155 Li-Kong St. Section 2,
              Peitou, Taipei, Taiwan 11221, Republic of China
              (E-mail:ympe@ym.edu.tw, URL:http://genome.ym.edu.tw/,
              Tel:886-2-28267043, Fax:886-2-28264930)
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 2 contigs. Gaps between the contigs
              * are represented as runs of N. The order of the pieces
              * is believed to be correct as given, however the sizes
              * of the gaps between them are based on estimates that have
              * provided by the submittor.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.

```

```

FEATURES
  source      1
              50664 50763: gap of 100 bp
              50764 100472: contig of 49709 bp in length.
              Location/Qualifiers
                1. 100472

```

```

BASE COUNT      30972 a 19378 c 19589 g 30429 t 104 others
ORIGIN

```

```

Query Match      81.7%: Score 18.8; DB 2; length 100472;
Best Local Similarity 90.9%: Pred. No.1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 AGTGGTGAATGTTGAGAGACCT 23
        |||||
Db      8050 AGTGGTGAATGTTGAGAGACCT 8071

```

```

Search completed: January 25, 2003, 04:29:33
Job time : 160.898 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 14:02:32 ; Search time 2252.1 Seconds

(without alignments)
11764.270 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctggtgaccagagaca.....ctctagctctgaccatg 8995

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11:  gb_sts:*
12:  gb_sy:*
13:  gb_un:*
14:  gb_vi:*
15:  em_ba:*
16:  em_fun:*
17:  em_hum:*
18:  em_in:*
19:  em_mu:*
20:  em_om:*
21:  em_or:*
22:  em_ov:*
23:  em_ph:*
24:  em_pl:*
25:  em_ro:*
26:  em_sts:*
27:  em_sy:*
28:  em_un:*
29:  em_vi:*
30:  em_htg_hum:*
31:  em_htg_inv:*
32:  em_htg_other:*
33:  em_htg_mus:*
34:  em_htg_pln:*
35:  em_htg_rtd:*
36:  em_htg_mam:*
37:  em_htg_vrt:*
38:  em_sy:*
39:  em_htgo_hum:*
40:  em_htgo_mus:*
41:  em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 8995 | 100.0 | 8995 | 6 AX027824 | AX027824 Sequence |
| 2 | 8762.2 | 97.4 | 192060 | 2 AC098570 | AC098570 Mus muscu |
| 3 | 3016.2 | 33.5 | 149109 | 2 AC101018 | AC101018 Rattus no |
| 4 | 938.6 | 10.4 | 159492 | 2 AC115173 | AC115173 Rattus no |
| 5 | 742.6 | 8.3 | 159492 | 2 AC115173 | AC115173 Rattus no |
| 6 | 188.6 | 2.1 | 80578 | 2 AC119514 | AC119514 Rattus no |
| 7 | 188.6 | 2.1 | 181772 | 2 AC094169 | AC094169 Rattus no |
| 8 | 188.6 | 2.1 | 194630 | 2 AC128412 | AC128412 Rattus no |
| 9 | 188 | 2.1 | 149425 | 2 AC013320 | AC013320 Homo sapi |
| 10 | 184.8 | 2.1 | 191754 | 9 AC021016 | AC021016 Homo sapi |
| 11 | 184.8 | 2.1 | 220909 | 2 AC115965 | AC115965 Mus muscu |
| 12 | 184 | 2.0 | 184591 | 2 AC102777 | AC102777 Mus muscu |
| 13 | 183.6 | 2.0 | 273800 | 10 AF100956 | AF100956 Mus muscu |
| 14 | 181.8 | 2.0 | 182695 | 2 AC130278 | AC130278 Mus muscu |
| 15 | 179.2 | 2.0 | 218502 | 2 AC025964 | AC025964 Mus muscu |
| 16 | 179.2 | 2.0 | 240536 | 10 AC025910 | AC025910 Mus muscu |
| 17 | 179.2 | 2.0 | 241400 | 10 AL646093 | AL646093 Mouse DNA |
| 18 | 178.2 | 2.0 | 198923 | 2 AC109263 | AC109263 Mus muscu |
| 19 | 177.8 | 2.0 | 165345 | 2 AC117095 | AC117095 Rattus no |
| 20 | 177.8 | 2.0 | 179805 | 2 AC122956 | AC122956 Rattus no |
| 21 | 177.6 | 2.0 | 224389 | 2 AC123811 | AC123811 Mus muscu |
| 22 | 177.4 | 2.0 | 208161 | 2 AC074145 | AC074145 Mus muscu |
| 23 | 177.2 | 2.0 | 165259 | 2 AL844206 | AL844206 Mus muscu |
| 24 | 177.2 | 2.0 | 166167 | 10 AL611936 | AL611936 Mouse DNA |
| 25 | 177 | 2.0 | 215049 | 2 AC122807 | AC122807 Mus muscu |
| 26 | 176.6 | 2.0 | 204573 | 10 AC087541 | AC087541 Mus muscu |
| 27 | 175.8 | 2.0 | 200057 | 2 AC073810 | AC073810 Mus muscu |
| 28 | 175 | 1.9 | 88410 | 2 AC098985 | AC098985 Rattus no |
| 29 | 174.4 | 1.9 | 207814 | 10 AL590994 | AL590994 Mouse DNA |
| 30 | 174.2 | 1.9 | 202095 | 2 AC122423 | AC122423 Mus muscu |
| 31 | 174 | 1.9 | 168280 | 2 AC123357 | AC123357 Rattus no |
| 32 | 173.6 | 1.9 | 119657 | 2 AC073667 | AC073667 Mus muscu |
| 33 | 173.6 | 1.9 | 170139 | 2 AC073727 | AC073727 Mus muscu |
| 34 | 172.8 | 1.9 | 249283 | 2 AC124577 | AC124577 Mus muscu |
| 35 | 172.4 | 1.9 | 73838 | 10 AL646020 | AL646020 Mouse DNA |
| 36 | 171.4 | 1.9 | 220242 | 10 AL603706 | AL603706 Mouse DNA |
| 37 | 171.2 | 1.9 | 186127 | 2 AC102630 | AC102630 Mus muscu |
| 38 | 170.6 | 1.9 | 159169 | 2 AC113976 | AC113976 Mus muscu |
| 39 | 170.4 | 1.9 | 188168 | 2 AC115806 | AC115806 Mus muscu |
| 40 | 170.4 | 1.9 | 195430 | 2 AC116140 | AC116140 Mus muscu |
| 41 | 170.4 | 1.9 | 207696 | 2 AC093473 | AC093473 Mus muscu |
| 42 | 170.2 | 1.9 | 135267 | 2 AC021434 | AC021434 Mus muscu |
| 43 | 170.2 | 1.9 | 158338 | 2 AL845258 | AL845258 Mus muscu |
| 44 | 170 | 1.9 | 166513 | 2 AC073787 | AC073787 Mus muscu |
| 45 | 170 | 1.9 | 182848 | 2 AC118017 | AC118017 Mus muscu |

ALIGNMENTS

```

RESULT 1
AX027824 LOCUS
DEFINITION Sequence 1 from Patent WO0034492.
ACCESSION AX027824
VERSION AX027824.1 GI:10188668
KEYWORDS
SOURCE
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 8995)
AUTHORS Robine,S., Louvard,D., Pinto,D. and Jaisser,F.
TITLE Regulatory sequences of the mouse villin gene - use in transgenesis
JOURNAL Patent: WO 0034492-A 1 15-JUN-2000;

```

ROBINE SYLVIE (FR) ; INST CURIE (FR) ; LOUYARD DANIEL (FR) ; PINTO DANIEL (FR) ; CENTRE NAT RECH SCIENT (FR) ; JAISSEY FREDERIC (FR)

FEATURES
Location/Qualifiers
1..8995

source
/organism="Mus sp."
/db_xref="taxon:10095"
3443..3487
exon
/note="exon 1"

Intron
BASE COUNT 2275 a 2105 c 2258 g 2357 t
ORIGIN

Query Match 100.0% Score 8995; DB 6; Length 8995;
Best Local Similarity 100.0% Pred. No. 0;
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTGTCACCAAGACACCTGTGTCCACACTGGGGAGTGGAGGAGAGGTGA 60
DB 1 GATCTGTCACCAAGACACCTGTGTCCACACTGGGGAGTGGAGGAGAGGTGA 60
QY 61 GAAGTTAAGTCATCTGTTTACATAGCAAGTTTCAGCCAGCTTACATGAAA 120
DB 61 GAAGTTAAGTCATCTGTTTACATAGCAAGTTTCAGCCAGCTTACATGAAA 120
QY 121 CCTTTGTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATAGAGGTTGG 180
DB 121 CCTTTGTTGTTGTTGTTGTTTAAAGCATTAATAATATACCATAGAGGTTGG 180
QY 181 CAGTGTGCAGACACCTTAATTCAGATTTCAGAGGACAGAGGACAGATCTCTGT 240
DB 181 CAGTGTGCAGACACCTTAATTCAGATTTCAGAGGACAGAGGACAGATCTCTGT 240
QY 241 GAGTTGGAAGTCAAGCTTACAGCAAGTCTAGTCCAGAGTGGCAAGGGCTACACAGAGA 300
DB 241 GAGTTGGAAGTCAAGCTTACAGCAAGTCTAGTCCAGAGTGGCAAGGGCTACACAGAGA 300
QY 301 AACCTGTCTCAATAAACCAGTACTAGTAGTAGTAATGCCATAGAGAAATTGGA 360
DB 301 AACCTGTCTCAATAAACCAGTACTAGTAGTAGTAATGCCATAGAGAAATTGGA 360
QY 361 GTCCATTTCAGATGACCATCTCTATAGATGATCTCTTGACCCAGCTAAGCTAATGTGA 420
DB 361 GTCCATTTCAGATGACCATCTCTATAGATGATCTCTTGACCCAGCTAAGCTAATGTGA 420
QY 421 TGGGGAAAGGGATGGGACTGTCTAGATTAAAGTGTGAGGGCATGCTATTCTCAA 480
DB 421 TGGGGAAAGGGATGGGACTGTCTAGATTAAAGTGTGAGGGCATGCTATTCTCAA 480
QY 481 TTTGATTCCATATGAAGAAGGCTGATTAAGGCCCAAGAGAAGTGGAACTGGAGCTTGACT 540
DB 481 TTTGATTCCATATGAAGAAGGCTGATTAAGGCCCAAGAGAAGTGGAACTGGAGCTTGACT 540
QY 541 GAAGACGTGACGGCTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCACAG 600
DB 541 GAAGACGTGACGGCTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCACAG 600
QY 601 CGTTACGTTTGAAGATCACTTTCAACACACAGAAAGAGTGGCTGCTGCTCAGC 660
DB 601 CGTTACGTTTGAAGATCACTTTCAACACACAGAAAGAGTGGCTGCTGCTCAGC 660
QY 661 GTAGCAGACACTGGCTGCAAGAGAGTATTTAGTGAAGAGTACCTTACATATCTTT 720
DB 661 GTAGCAGACACTGGCTGCAAGAGAGTATTTAGTGAAGAGTACCTTACATATCTTT 720
QY 721 GCACCTTATACATACACAGTGTCAATGTCTAACTCCTAGTCCACAGATGCTGTTTACA 780
DB 721 GCACCTTATACATACACAGTGTCAATGTCTAACTCCTAGTCCACAGATGCTGTTTACA 780
QY 781 CTCGTTTCGCTTTGCCATCTGTTGACATTTGTCAGAACCAAAATTTGGAATTTGGGT 840
DB 781 CTCGTTTCGCTTTGCCATCTGTTGACATTTGTCAGAACCAAAATTTGGAATTTGGGT 840
QY 841 ATTATTTTGTGCTGAGGACACCATTCAGGGCTTTTCACTTTCAGGCACATGTTTAC 900

DB 841 ATTATTTTGTGCTGAGGACACCATTCAGGGCTTTTCACTTTCAGGCACATGTTTAC 900
QY 901 TACAGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960
DB 901 TACAGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960
QY 961 GCATGAGGTAGCATGTATACGTATGTATAGAGTCATGATGAGTGGCTACCTCAAA 1020
DB 961 GCATGAGGTAGCATGTATACGTATGTATAGAGTCATGATGAGTGGCTACCTCAAA 1020
QY 1021 ATCATTGACAGATCCCCAGCAAGTGAAGTACCAGGCTTTGATGTTGTTATGAGGACTG 1080
DB 1021 ATCATTGACAGATCCCCAGCAAGTGAAGTACCAGGCTTTGATGTTGTTATGAGGACTG 1080
QY 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGTGGCTTAACCATGGACACAGCTCT 1140
DB 1081 GGAGCCAAAGGCTGGGTTCTCTGCAAGAGCAGCAGTGGCTTAACCATGGACACAGCTCT 1140
QY 1141 CTAGGCTTAAGTAACTTTTAAATATATATTCACAGCCGGGTGTGTGTC 1200
DB 1141 CTAGGCTTAAGTAACTTTTAAATATATATTCACAGCCGGGTGTGTGTC 1200
QY 1201 ACAGCCCTTTAATCCAGCACTTGAGAGGCTGAGTGTAGCAATTATACACACAGCCAG 1260
DB 1201 ACAGCCCTTTAATCCAGCACTTGAGAGGCTGAGTGTAGCAATTATACACACAGCCAG 1260
QY 1261 CTGGGGTGCAGAGCTGGGCTGTTTGTGTTTGTGTTTCTTATATGTCATGCTGTCTTA 1320
DB 1261 CTGGGGTGCAGAGCTGGGCTGTTTGTGTTTGTGTTTCTTATATGTCATGCTGTCTTA 1320
QY 1321 CCTGCTGATGTCCGTGCAAGGGTGTAGATCCCTTGAGAGCTGAGTTAAAGACAGTTG 1380
DB 1321 CCTGCTGATGTCCGTGCAAGGGTGTAGATCCCTTGAGAGCTGAGTTAAAGACAGTTG 1380
QY 1381 TGATCAGCTGCCGTTTACAGATGCTGGAATTTGAACCCAGGCTGCTTAAGAAACAGCC 1440
DB 1381 TGATCAGCTGCCGTTTACAGATGCTGGAATTTGAACCCAGGCTGCTTAAGAAACAGCC 1440
QY 1441 AGTGTCTTAATCTTGAGCCACCCCTCCACACCGCTTTTGAAGACCTTAACCTTTTG 1500
DB 1441 AGTGTCTTAATCTTGAGCCACCCCTCCACACCGCTTTTGAAGACCTTAACCTTTTG 1500
QY 1501 TGTAAATGTGGAACTGAGTGTGCTTGCACCTTACCAAGTGTGCTGCTTACATCA 1560
DB 1501 TGTAAATGTGGAACTGAGTGTGCTTGCACCTTACCAAGTGTGCTGCTTACATCA 1560
QY 1561 CTGAGCCGCTTACCCACACAGCACTAGTGTATACAGTTTAAGGGCAACACTTAACAATGACA 1620
DB 1561 CTGAGCCGCTTACCCACACAGCACTAGTGTATACAGTTTAAGGGCAACACTTAACAATGACA 1620
QY 1621 ATAGTTGGATTAAGTTTGAATATAGTCCCTGAGACCTATGTTGCTGAGCCCTTGTGCTCC 1680
DB 1621 ATAGTTGGATTAAGTTTGAATATAGTCCCTGAGACCTATGTTGCTGAGCCCTTGTGCTCC 1680
QY 1681 TTAGCATGTGCTGTGAGAAATGAGAAAAAGAACTGAGTGTGCTGAGAACCCACA 1740
DB 1681 TTAGCATGTGCTGTGAGAAATGAGAAAAAGAACTGAGTGTGCTGAGAACCCACA 1740
QY 1741 GAGGACGCGGAGAACCCACTCTGAAAGTTGTTCTGAGCTTCAATCAACACTTCAAT 1800
DB 1741 GAGGACGCGGAGAACCCACTCTGAAAGTTGTTCTGAGCTTCAATCAACACTTCAAT 1800
QY 1801 AATAGTTAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
DB 1801 AATAGTTAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
QY 1861 GAGATGCTCAGCTTCCAGAGACACTGTGCTCTTGGAGAGGACTTACATTCAGTTCC 1920
DB 1861 GAGATGCTCAGCTTCCAGAGACACTGTGCTCTTGGAGAGGACTTACATTCAGTTCC 1920
QY 1921 AGGACTCATATGTTGGCTCAGACCATCTGTAATTCACAGTTCCAGAGGCTTCCACACCT 1980

Dh 1921 AGGATCATATGTGGCTACAGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCCT 1980
Qy 1981 CTTTGCGCTCCACAGGACACCACATACATAGTACAGACATACATAGTACAGGCAAAACACC 2040
Dh 1981 CTTTGCGCTCCACAGGACACCACATACATAGTACAGACATACATAGTACAGGCAAAACACC 2040
Qy 2041 CATACACACATAAATTAATTAAGAACTTAAAGAGTGCATGTGTTGGTAAACATTTGCT 2100
Dh 2041 CATACACACATAAATTAATTAAGAACTTAAAGAGTGCATGTGTTGGTAAACATTTGCT 2100
Qy 2101 TACACATGTGATTTGAAGACATGTACAAACGACACACTGAAGAGGATCTGGGCTGGAG 2160
Dh 2101 TACACATGTGATTTGAAGACATGTACAAACGACACACTGAAGAGGATCTGGGCTGGAG 2160
Qy 2161 AGATGGCTGACGGGTTAAGACACTGCTCTCCGAAGAGAGTCTTGAGTTCAAT 2220
Dh 2161 AGATGGCTGACGGGTTAAGACACTGCTCTCCGAAGAGAGTCTTGAGTTCAAT 2220
Qy 2221 CCTAGCAACACATGTGGGCTCAGACACCATCATATGAGTCTGACACCCTTTGCT 2280
Dh 2221 CCTAGCAACACATGTGGGCTCAGACACCATCATATGAGTCTGACACCCTTTGCT 2280
Qy 2281 GCATCTGAGACAGCTCGACAGCTACAGTGTACTTAAATTAATTAATTAATTAATTTT 2340
Dh 2281 GCATCTGAGACAGCTCGACAGCTACAGTGTACTTAAATTAATTAATTAATTAATTTT 2340
Qy 2341 TTTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGTTTATGAGCATGACTCAGC 2400
Dh 2341 TTTTAAAAAATGAGAGGATCTGACACCTCAAAAGAGTTTATGAGCATGACTCAGC 2400
Qy 2401 GGTGATTTATCTATCTGAGAGTTTCTTCCGCTTGCTGCAACTGGGAGGAGCAGAC 2460
Dh 2401 GGTGATTTATCTATCTGAGAGTTTCTTCCGCTTGCTGCAACTGGGAGGAGCAGAC 2460
Qy 2461 CCCCTTTTCATTCACAGAAAGCGGTGCTACATTAATTTCTGAACAAAACAGCAGCTGAGT 2520
Dh 2461 CCCCTTTTCATTCACAGAAAGCGGTGCTACATTAATTTCTGAACAAAACAGCAGCTGAGT 2520
Qy 2521 ATGTTTACTGCTCTGCTGATATGAGACGCGGACGCGGCGGCGACACACACACAC 2580
Dh 2521 ATGTTTACTGCTCTGCTGATATGAGACGCGGACGCGGCGGCGACACACACACAC 2580
Qy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGAAAGTCA 2640
Dh 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGAAAGTCA 2640
Qy 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTAAAGAGACACAGATTCC 2700
Dh 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTAAAGAGACACAGATTCC 2700
Qy 2701 AAGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2760
Dh 2701 AAGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2760
Qy 2761 AGAGACCTACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820
Dh 2761 AGAGACCTACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2820
Qy 2821 CCCCTTTGTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
Dh 2821 CCCCTTTGTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
Qy 2881 GCTCATTAAGAGTGTGAGGCTGCTCACTCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2940
Dh 2881 GCTCATTAAGAGTGTGAGGCTGCTCACTCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 2940
Qy 2941 GCCCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000
Dh 2941 GCCCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000
Qy 3001 CTGTCTGTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3060
Dh 3001 CTGTCTGTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3060

Qy 3061 CAAGACCCCAAGAGTCTCTACATCCCATCCAGTGTGCCCCCTGCCCCGACACACC 3120
Dh 3061 CAAGACCCCAAGAGTCTCTACATCCCATCCAGTGTGCCCCCTGCCCCGACACACC 3120
Qy 3121 CCCCCGACTCCCGTGCACATCTCTAGAGGCTGAGGGGTGGGACCCCTGTGTGGGGTTGC 3180
Dh 3121 CCCCCGACTCCCGTGCACATCTCTAGAGGCTGAGGGGTGGGACCCCTGTGTGGGGTTGC 3180
Qy 3181 CTACCTGAGGTAGAGCCACAGTCTTAGCCGGAAGTCCACCCATCCCTGAAGCTGAGA 3240
Dh 3181 CTACCTGAGGTAGAGCCACAGTCTTAGCCGGAAGTCCACCCATCCCTGAAGCTGAGA 3240
Qy 3241 GCCAAGGGGGGGGACACGGGAGCTGAGGCTGTGAGGCTGTGCTGGGCTCTAGTTCC 3300
Dh 3241 GCCAAGGGGGGGGACACGGGAGCTGAGGCTGTGAGGCTGTGCTGGGCTCTAGTTCC 3300
Qy 3301 AGGACCTGGGACATCTCTCCACCCCCCATCTCTCTGCTGGGAGCTATCTTC 3360
Dh 3301 AGGACCTGGGACATCTCTCCACCCCCCATCTCTCTGCTGGGAGCTATCTTC 3360
Qy 3361 CTTATATGTGAAGAGTCTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3420
Dh 3361 CTTATATGTGAAGAGTCTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3420
Qy 3421 GTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3480
Dh 3421 GTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3480
Qy 3481 CAGACAGGTAAAGCAATTTGGTGGGAGACATAGTGAACAGAGTGGTTGAGGGAGAG 3540
Dh 3481 CAGACAGGTAAAGCAATTTGGTGGGAGACATAGTGAACAGAGTGGTTGAGGGAGAG 3540
Qy 3541 GGTCTTGTCTCTCTCTGAGAGCTGTGCTTCTGTAAGACCTTGGTAAGTTTGGGGG 3600
Dh 3541 GGTCTTGTCTCTCTCTGAGAGCTGTGCTTCTGTAAGACCTTGGTAAGTTTGGGGG 3600
Qy 3601 TGAAGTAAAGTGTCTGAAGTCTGAAGAAAGCAAGAGGAGGAGGAGGAGGAGGAGG 3660
Dh 3601 TGAAGTAAAGTGTCTGAAGTCTGAAGAAAGCAAGAGGAGGAGGAGGAGGAGGAGG 3660
Qy 3661 TCAATGAAGAAAGTTCACAGACCCCTTCTCTGAAGTCACTTCCCTGATCTGTAG 3720
Dh 3661 TCAATGAAGAAAGTTCACAGACCCCTTCTCTGAAGTCACTTCCCTGATCTGTAG 3720
Qy 3721 ATTCCCTGGGACCAAGGTGCTCTGAGACTCAGATTCTAACAATCAAGAGAGT 3780
Dh 3721 ATTCCCTGGGACCAAGGTGCTCTGAGACTCAGATTCTAACAATCAAGAGAGT 3780
Qy 3781 CTTGAGACTTGGAGTCCGCTGCTATTTACTTCTCTGCTGCTGCTGCTGCTGCTGCT 3840
Dh 3781 CTTGAGACTTGGAGTCCGCTGCTGCTATTTACTTCTCTGCTGCTGCTGCTGCTGCT 3840
Qy 3841 TCAATGCTTACATCTGAAGTGTCTTGTGTACATTTCCCTGACACTCTGGGA 3900
Dh 3841 TCAATGCTTACATCTGAAGTGTCTTGTGTACATTTCCCTGACACTCTGGGA 3900
Qy 3901 GGTGATTCCTTGGACATGTATCTGAGTGTGAAGCTGACGCCAGCAGAGAGAGGGG 3960
Dh 3901 GGTGATTCCTTGGACATGTATCTGAGTGTGAAGCTGACGCCAGCAGAGAGAGGGG 3960
Qy 3961 AGAGTCAAGAGTGTGTCTAGGCTCTAATTAAGGCTGTGAACATCACCCCTTCTGAAG 4020
Dh 3961 AGAGTCAAGAGTGTGTCTAGGCTCTAATTAAGGCTGTGAACATCACCCCTTCTGAAG 4020
Qy 4021 GGGCCCTCAATTTTTCGTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGGCA 4080
Dh 4021 GGGCCCTCAATTTTTCGTTACCATGATCTATTTTATATCAGAGTGGGAGTGAAGGCA 4080
Qy 4081 AACCTGCCAAGATTGGAGTCACTCAGACCAAGGTAATCTGCTCAGAAATCCCTG 4140
Dh 4081 AACCTGCCAAGATTGGAGTCACTCAGACCAAGGTAATCTGCTCAGAAATCCCTG 4140

| | | | | |
|----|------|-------|--|------|
| QY | 4141 | TCAC | TTGAGGTTGGGAGAAATCTCCCTCTGGGGGGCTTCCAGGCTTGGTTAGCAGAGAGGT | 4200 |
| Db | 4141 | TCAC | TTTGGAGGTTGGGAGAAATCTCCCTCTGGGGGGCTTCCAGGCTTGGTTAGCAGAGAGGT | 4200 |
| QY | 4201 | ATCC | TTGTATAGGGCATGTACCTAGCTATAGGGGTTACTACATTCCTCTGCCAGTTAAAG | 4260 |
| Db | 4201 | ATCC | TTTGTATAGGGCATGTACCTAGCTATAGGGGTTACTACATTCCTCTGCCAGTTAAAG | 4260 |
| QY | 4261 | CTGG | AACTAAAACCCACGCGCGCCAGGATTTCTTCTACAGTTGTACCAGAAACACA | 4320 |
| Db | 4261 | CTGG | AACTAAAACCCACGCGCGCCAGGATTTCTTCTACAGTTGTACCAGAAACACA | 4320 |
| QY | 4321 | AGACA | GTAGATATGCAAGGATAGCTAGCTGGGGAGAGAAGAACTTAAACCCCCCAAG | 4380 |
| Db | 4321 | AGACA | GTAGATATGCAAGGATAGCTAGCTGGGGAGAGAAGAACTTAAACCCCCCAAG | 4380 |
| QY | 4381 | GCC | CAAGGTTCCGTTCCCTAGTTTCCAAATGCGAGTATGAGTGCATAGCTATAGGGGTG | 4440 |
| Db | 4381 | GCC | CAAGGTTCCGTTCCCTAGTTTCCAAATGCGAGTATGAGTGCATAGCTATAGGGGTG | 4440 |
| QY | 4441 | TGAG | TTGGTGGTACGATCAAGCATGAGATGTGCTCATGCTGTAGTGTGTAAATCTGACAC | 4500 |
| Db | 4441 | TGAG | TTGGTGGTGGTACGATCAAGCATGAGATGTGCTCATGCTGTGTAAATCTGACAC | 4500 |
| QY | 4501 | TTGG | GAGGCTGCAAGCAGAGAGATTTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA | 4560 |
| Db | 4501 | TTGG | GAGGAGGCTGCAAGCAGAGAGATTTGCTATATGTTTGAGGCCAGCTGAGCTATAGAGCA | 4560 |
| QY | 4561 | GACT | TTTGTCTTTAAGAAAAATGAAAGCCCGACAGTGGTGGCACAGCCCTTTAATCCA | 4620 |
| Db | 4561 | GACT | TTTGTGTCTTTAAGAAAAATGAAAGCCCGACAGTGGTGGCACAGCCCTTTAATCCA | 4620 |
| QY | 4621 | GCAC | TTGGGAGGCGAGAGCAGGAGATTTTCTGAGTTCAAGGCCACCTGGTCTATAGAGT | 4680 |
| Db | 4621 | GCAC | TTTGGGAGGCGAGAGCAGGAGATTTTCTGAGTTCAAGGCCACCTGGTCTATAGAGT | 4680 |
| QY | 4681 | GAGT | TCAGGACGACGCGGCTACACAGAGAAACCCCTGTTTGAAGAAACCAAGAAAAACA | 4740 |
| Db | 4681 | GAGT | TCAGTTCCAGGACGCGGCTACACAGAGAAACCCCTGTTTGAAGAAACCAAGAAAAACA | 4740 |
| QY | 4741 | AACA | AAAAACAAACAAACCAAAACCCCAACCCCAACCTCTCATCTCTCATCTCTC | 4800 |
| Db | 4741 | AACA | AAAAACAAACAAACCAAAACCCCAACCCCAACCCCAACCTCTCATCTCTCTCTC | 4800 |
| QY | 4801 | TAG | CTGTGCTGTCTGATAGTGTAGATTTGGGAGCTTGAGCTTATATATTAATAGGCC | 4860 |
| Db | 4801 | TAG | CTGTGTCTGTCTGATAGTGTAGATTTGGGAGCTTGAGCTTATATATTAATAGGCC | 4860 |
| QY | 4861 | TTTT | TTATCACTGGTCAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTTGAGA | 4920 |
| Db | 4861 | TTTT | TTTATTCACCTGGTCAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTTGAGA | 4920 |
| QY | 4921 | AAG | TATCTTGGCCGACCAAAATTTCTGGGAAGCTTCTGAGAGAAAGTGTGCCAT | 4980 |
| Db | 4921 | AAG | TATCTTGGCCGACCAAAATTTCTGGGAAGCTTCTGAGAGAAAGTGTGCCAT | 4980 |
| QY | 4981 | CAGA | CTACTGTCTTCAAGACGAGAGAGAGGTTGGAATAGTTGTGTGACAGACAGTT | 5040 |
| Db | 4981 | CAGA | CTACTGTCTTCAAGAGGCAAGAGAGAGGTTGGAATAGTTGTGTGTGACAGACAGTT | 5040 |
| QY | 5041 | GGA | CACAGAGACAGAGAGGGGAGGATCCACAGATTTCTGACATGTAGCTGACTTTTGGT | 5100 |
| Db | 5041 | GGA | CACACAGAGAGACAGAGAGGGGAGGATCCACAGATTTCTGACATGTAGCTGACTTTTGGT | 5100 |
| QY | 5101 | TCT | CTGGGTGACAGTGTCCCCCAGGGATATAGGCTTTAGAAAAGGGGACCAAGGGTGAACC | 5160 |
| Db | 5101 | TCT | CTGGGTGTACAAGTGTCTCCCAAGGATATAGGCTTTAGAAAAGGGGACCAAGGGTGAACC | 5160 |
| QY | 5161 | AAT | GATTCAGGTTGAGGGACACATCAGGCCAGGGTCTCTGCTGGCAAGCTTAAAGATG | 5220 |
| Db | 5161 | AAT | GATGTTCAGGTTGAGGGACACATCAGGCCAGGGTCTCTGCTGGCAAGCTTAAAGATG | 5220 |
| QY | 5221 | AGAG | CCCTTAAACCTCTCTGAGATTTTAGGGGAGACAGAGAGAGCTGAGAGATCTCTTCTA | 5280 |

| | | | | |
|----|------|---|--|------|
| Dd | 5221 | AGAGCCCTTACACCTCCCTGAAAGTTAGGGGAGACAGAGAGCTGAGGAGATCTTCTA | | 5280 |
| Qy | 5281 | GGGTGAAGAGAGGTATCTCTCTGACAACAATGAGCTAGAGACAGAAAGCATTTGGACAG | | 5340 |
| Dd | 5281 | GGGTGAAGAGAGGTATCTCTCTGACAACAATGAGCTAGAGACAGAAAGCATTTGGACAG | | 5340 |
| Qy | 5341 | TTAACCCCTCAGAACACGACATCCCTCTCTGCGCTAAGAGAGCTGGGCCCTTCTGTTT | | 5400 |
| Dd | 5341 | TTAACCCCTCAGAACACGACATCCCTCTCTGCGCTAAGAGAGCTGGGCCCTTCTGTTT | | 5400 |
| Qy | 5401 | AAGAACTTACTTTTCTTCAGAGAGAGGACAGAGCTTTGTCCTCTCTGTTGGTCA | | 5460 |
| Dd | 5401 | AAGAACTTACTTTTCTTCAGAGAGAGGACAGAGCTTTGTCCTCTCTGTTGGTCA | | 5460 |
| Qy | 5461 | TAAACACCCCGTGTGTAAACATTAAGTTATTTACTGTGAGTTTGCTCCAGACAGTCCA | | 5520 |
| Dd | 5461 | TAAACACCCCGTGTGTAAACATTAAGTTATTTACTGTGAGTTTGCTCCAGACAGTCCA | | 5520 |
| Qy | 5521 | TCTGGTAGACCTGCTGCTTAAGTCAACAGATATGGCCACATTCCTTACCAGAAAG | | 5580 |
| Dd | 5521 | TCTGGTAGACCTGCTGCTTAAGTCAACAGATATGGCCACATTCCTTACCAGAAAG | | 5580 |
| Qy | 5581 | TGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACATGCGCAGATTAACAAAAA | | 5640 |
| Dd | 5581 | TGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACATGCGCAGATTAACAAAAA | | 5640 |
| Qy | 5641 | ACTACTATCCCTTTTACCACAAATTTGGTTTTGCTGAACCCAGAGAGGGGTGTGAGTGTAT | | 5700 |
| Dd | 5641 | ACTACTATCCCTTTTACCACAAATTTGGTTTTGCTGAACCCAGAGAGGGGTGTGAGTGTAT | | 5700 |
| Qy | 5701 | GTTGT | | 5760 |
| Dd | 5701 | GTTGT | | 5760 |
| Qy | 5761 | CTTGGGGACTTTTCATGCTAAAGATATCTGATATTGCGGCCATGCCAACAGGGTAT | | 5820 |
| Dd | 5761 | CTTGGGGACTTTTCATGCTAAAGATATCTGATATTGCGGCCATGCCAACAGGGTAT | | 5820 |
| Qy | 5821 | TGGGGAGAGTCAGGCTTCTGCAAAACAGTAAACCTGCCAGAGTGGATTGGTGGCTGAA | | 5880 |
| Dd | 5821 | TGGGGAGAGTCAGGCTTCTGCAAAACAGTAAACCTGCCAGAGTGGATTGGTGGCTGAA | | 5880 |
| Qy | 5881 | TCACCAAGGGGACAGCTGATCAGAGTGTGACAGAACTCAACAAGATAAGCCACCTGTGGG | | 5940 |
| Dd | 5881 | TCACCAAGGGGACAGCTGATCAGAGTGTGACAGAACTCAACAAGATAAGCCACCTGTGGG | | 5940 |
| Qy | 5941 | GCTCAGAAAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTATTATTCGAAGACATGAC | | 6000 |
| Dd | 5941 | GCTCAGAAAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTATTATTCGAAGACATGAC | | 6000 |
| Qy | 6001 | TCAAAATCAAGGTCAAGAGAGATTAGCTGAGAGATGGGGCTGCACGTGTGGGACACC | | 6060 |
| Dd | 6001 | TCAAAATCAAGGTCAAGAGAGATTAGCTGAGAGATGGGGCTGCACGTGTGGGACACC | | 6060 |
| Qy | 6061 | TGACCTTGCACCTTATTTAGTCACTAAGGCCAAGAGACGTACAGAGAGGTACTGGGTCTA | | 6120 |
| Dd | 6061 | TGACCTTGCACCTTATTTAGTCACTAAGGCCAAGAGACGTACAGAGAGGTACTGGGTCTA | | 6120 |
| Qy | 6121 | CTCAGCTTGGAGCAGGACGTGGAGATGGGTACTCTCATCTCTGATGGAGAGGCTGAG | | 6180 |
| Dd | 6121 | CTCAGCTTGGAGCAGGACGTGGAGATGGGTACTCTCATCTCTGATGGAGAGGCTGAG | | 6180 |
| Qy | 6181 | CACCAACAGGTACAAAGTTCCTCGTGTCTCATAGCCAGGATTCCTGGCCAGTTTCAAG | | 6240 |
| Dd | 6181 | CACCAACAGGTACAAAGTTCCTCGTGTCTCATAGCCAGGATTCCTGGCCAGTTTCAAG | | 6240 |
| Qy | 6241 | GACTAAGAGACTCATCTGTGTGGAAACAAGTATCCAAAGCCCTAAGCCCCATTTTGGTCT | | 6300 |
| Dd | 6241 | GACTAAGAGACTCATCTGTGTGGAAACAAGTATCCAAAGCCCTAAGCCCCATTTTGGTCT | | 6300 |
| Qy | 6301 | AATTAATCAGAACCCCTGGGGATGACAGCTGTGACAGCAGAGCTTTTAAAAAGCTC | | 6360 |

Db 6301 AATTAAATCAGAACCCCTGGGGATGCAAGCTCTGAGCAGCAGAGACTTTTAAAAAGCTC 6360
Qy 6361 CCAGGTGATTCGATACAGACAGCTGGAAACAACAGACGTACAGGTTCACAACGAAGAAGGC 6420
Db 6361 CCAGGTGATTCGATACAGACAGCTGGAAACAACAGACGTACAGGTTCACAACGAAGAAGGC 6420
Qy 6421 AAAGCTAGGGAAGAGTGGGATGGGAGACCTTCTTCCAGGCGAGTAGATGAGGCTGGTT 6480
Db 6421 AAAGCTAGGGAAGAGTGGGATGGGAGACCTTCTTCCAGGCGAGTAGATGAGGCTGGTT 6480
Qy 6481 AGCAGGTGGCAGCCTTCCTCTGCTGCTGTCATATAGCTATATCCATCCATCCATCCAT 6540
Db 6481 AGCAGGTGGCAGCCTTCCTCTGCTGCTGTCATATAGCTATCCATCCATCCATCCAT 6540
Qy 6541 ACACCCACCCATCCATTTATATGACCCCATCCCTTCCATCCATCCATCCATCCATCCAC 6600
Db 6541 ACACCCACCCATCCATTTATGACCCCATCCCTTCCATCCATCCATCCATCCATCCAC 6600
Qy 6601 CCAGGCATCCATCCAAACCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCAGGCATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Qy 6661 CATTATTCACACAGAGAACTGGTATTTGACTAAATGGGAGATTAAATTAATTTTGA 6720
Db 6661 CATTATTCACACAGAGAACTGGTATTTGACTAAATGGGAGATTAAATTAATTTTGA 6720
Qy 6721 AGCTCTGTGATTCAGCTGATTTGTGATGTATGTGACAGGTACATACACAGCACAGCTG 6780
Db 6721 AGCTCTGTGATTCAGCTGATTTGTGATGTATGTGACAGGTACATACACAGCACAGCTG 6780
Qy 6781 TGGCAATCGAGAAAGGTTTTGGGGTGTGTTTTCTTCCACCGGTGGGTTCTGGGGA 6840
Db 6781 TGGCAATCGAGAAAGGTTTTGGGGTGTGTTTTCTTCCACCGGTGGGTTCTGGGGA 6840
Qy 6841 TTGAACCTCAAAATTAATCGGGCTGGTGGCAAGTGTCTTTACACCGAGCCATTGGCTGACA 6900
Db 6841 TTGAACCTCAAAATTAATCGGGCTGGTGGCAAGTGTCTTTACACCGAGCCATTGGCTGACA 6900
Qy 6901 CATCATTTATTTTAGAAAGCATCTTATGTATGTACAGGCTGGCCTTCAGCTTGTATGTG 6960
Db 6901 CATCATTTATTTTAGAAAGCATCTTATGTATGTACAGGCTGGCCTTCAGCTTGTATGTG 6960
Qy 6961 CCACGAGTGAATGCTTTAATCCTGCTTCCAGCCCTCCAGGAGTGTAGGTTTACAGGT 7020
Db 6961 CCACGAGTGAATGCTTTAATCCTGCTTCCAGCCCTCCAGGAGTGTAGGTTTACAGGT 7020
Qy 7021 GTTCAACTGTGTAATGCTTTAATCCAGCACTCTGTGGGGGGGGGGGAGGCGGATC 7080
Db 7021 GTTCAACTGTGTAATGCTTTAATCCAGCACTCTGTGGGGGGGGGGGAGGCGGATC 7080
Qy 7081 CCTGAGTTGGAGGCGAGTTTGGTCTACAGAGTTTCAGAGTACCTTGGGGCTATACAGGAA 7140
Db 7081 CCTGAGTTGGAGGCGAGTTTGGTCTACAGAGTTTCAGAGTACCTTGGGGCTATACAGGAA 7140
Qy 7141 ACCCTATCCCAAAACAACAAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200
Db 7141 ACCCTATCCCAAAACAACAAACAACAACAACAACAACAACAACAACAACAACAACAACA 7200
Qy 7201 TTAGAGATATTTAGAGGTAGTAGGGCTGTGAGGGAGAGTCAATGCTTCTTTTGTATTT 7260
Db 7201 TTAGAGATATTTAGAGGTAGTAGGGCTGTGAGGGAGAGTCAATGCTTCTTTTGTATTT 7260
Qy 7261 ATATATAGTAAAGTACTCACAAGATGATTTATCTATCTATCTATCTATCTATCTATCT 7320
Db 7261 ATATATAGTAAAGTACTCACAAGATGATTTATCTATCTATCTATCTATCTATCTATCT 7320
Qy 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7380
Qy 7381 CTGCTTTGACTCTGAATGCTCTATTTCTGTGGGTCAACTCTTCCACCCCTAGTGGGTTT 7440
Db 7381 CTGCTTTGACTCTGAATGCTCTATTTCTGTGGGTCAACTCTTCCACCCCTAGTGGGTTT 7440

Qy 7441 ACCAAGACCCAGACATTTATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 7500
Db 7441 ACCAAGACCCAGACATTTATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 7500
Qy 7501 GGACTCAGGGTCTTGTGCACTCTTAAGCAAGCTCTGCGACAGAGCTGCACCTCCAGTCC 7560
Db 7501 GGACTCAGGGTCTTGTGCACTCTTAAGCAAGCTCTGCGACAGAGCTGCACCTCCAGTCC 7560
Qy 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTCAATATTCGACGGCTATGTAGCTCTCTC 7620
Db 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTCAATATTCGACGGCTATGTAGCTCTCTC 7620
Qy 7621 CACCTCCAGTTCCAGACACTTCTGTCATCCAGTGGGGGGGGAACCTCTGTCACACA 7680
Db 7621 CACCTCCAGTTCCAGACACTTCTGTCATCCAGTGGGGGGGGAACCTCTGTCACACA 7680
Qy 7681 GTGCCCTGTCCCTGTCTTACAGCTACATATTTGCGCTGTGAACAGTTCATGTAAATG 7740
Db 7681 GTGCCCTGTCCCTGTCTTACAGCTACATATTTGCGCTGTGAACAGTTCATGTAAATG 7740
Qy 7741 GGATGCGTTCGTGTATCTTTTATGGCTGGCCCTTTATCTTACACAGATTTGCTGTG 7800
Db 7741 GGATGCGTTCGTGTATCTTTTATGGCTGGCCCTTTATCTTACACAGATTTGCTGTG 7800
Qy 7801 GGCCATGTGTACGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860
Db 7801 GGCCATGTGTACGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860
Qy 7861 TGTGTGATTAACACATTTCTGTTTCATTTACTGATGGAATTTGTGGCCCCCACCAC 7920
Db 7861 TGTGTGATTAACACATTTCTGTTTCATTTACTGATGGAATTTGTGGCCCCCACCAC 7920
Qy 7921 CTTTTTTTTTTTTTATTTGAGACAAGTCTTCTGTATCTTGTGATCTTGTGCTGTCT 7980
Db 7921 CTTTTTTTTTTTTTATTTGAGACAAGTCTTCTGTATCTTGTGATCTTGTGCTGTCT 7980
Qy 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGAACCTCTGTGAAC 8040
Db 7981 GAGCTCACTGTGTAGACAGGCTGTGAGGCTGTCTTCCACTTTTGAACCTCTGTGTGAAC 8040
Qy 8041 AGAGTAGCAGTGAACCTTCAAGACAAATTTCTGTTTTGTTTTGTTTTTATCATTTGTGTG 8100
Db 8041 AGAGTAGCAGTGAACCTTCAAGACAAATTTCTGTTTTGTTTTGTTTTTATCATTTGTGTG 8100
Qy 8101 TGTATGCGTGTATATGTGATGATTTGTGTCTTCAAGTGTGTCACATGTGTGATCTGTGTG 8160
Db 8101 TGTATGCGTGTATATGTGATGATTTGTGTCTTCAAGTGTGTCACATGTGTGATCTGTGTG 8160
Qy 8161 TGGGACAGAGAAACAACCGAGTGTCCATCTCCAGATACATGCAATCTTGTAAATATGTA 8220
Db 8161 TGGGACAGAGAAACAACCGAGTGTCCATCTCCAGATACATGCAATCTTGTAAATATGTA 8220
Qy 8221 TGTATATATGTATGTATTTATTTAGTGTGCGCAAGTATGCAAGTATTTTGTGGAGTTTTT 8280
Db 8221 TGTATATATGTATGTATTTATTTAGTGTGCGCAAGTATGCAAGTATTTTGTGGAGTTTTT 8280
Qy 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTCAGCTCTGCGGCTATGAGCAATGCTCTTCA 8340
Db 8281 CTTCCTTTGTGGGCTCTCCGCAATTAACCTCAGCTCTGCGGCTATGAGCAATGCTCTTCA 8340
Qy 8341 CTGATGAGCCATCTGCTGTGCGCTGCTGCGCACTCTCTCTTATTTTCCCAATGGGACTA 8400
Db 8341 CTGATGAGCCATCTGCTGTGCGCTGCTGCGCACTCTCTCTTATTTTCCCAATGGGACTA 8400
Qy 8401 CGCAGTCACTGGGCTTAAAGCTTCAACAAGTCAATGAGTGGTAAAGCAGGAGACTACAG 8460
Db 8401 CGCAGTCACTGGGCTTAAAGCTTCAACAAGTCAATGAGTGGTAAAGCAGGAGACTACAG 8460
Qy 8461 GGATATGCTGGCTCTGCTGCTCAGAGTGTGAATTAACAGGCATACATCACTGCTGGAAG 8520
Db 8461 GGATATGCTGGCTCTGCTGCTCAGAGTGTGAATTAACAGGCATACATCACTGCTGGAAG 8520

| | | | |
|----|------|---|------|
| QY | 8521 | ATTTTAACTCGAATCCCTGAGATGTGACACAGCACTTCAACAATGAGAGGTTCTTTTGT | 8580 |
| Db | 8521 | ATTTTAACTCGAATCCCTGAGATGTGACACAGCACTTCAACAATGAGAGGTTCTTTTGT | 8580 |
| QY | 8561 | GTTTGGCTTTGGCTTCTCTGCGATAAGATCAGGACAGCTGGAATAGTGTAGGCTGGGCTAC | 8640 |
| Db | 8561 | GTTTGGCTTTGGCTTCTCTGCGATAAGATCAGGACAGCTGGAATAGTGTAGGCTGGGCTAC | 8640 |
| QY | 8641 | ATAACATCTTCTTCACAAAGCCCTTATAGGTATGAGGAGCTCGAGGCTTAAGCAAGAGCCTT | 8700 |
| Db | 8641 | ATAACATCTTCTTCACAAAGCCCTTATAGGTATGAGGAGGAGCTCGAGGCTTAAGCAAGAGCCTT | 8700 |
| QY | 8701 | AAGCGGCTGTGATAGCACACAGAGTATAGCTGCACTATATATAGCAAGACCTTGTTCAAAA | 8760 |
| Db | 8701 | AAGCGGCTGTGATAGCACACAGAGTATAGCTGCACTATATATAGCAAGACCTTGTTCAAAA | 8760 |
| QY | 8761 | ACATGAGGAGGAGGGGTATGTTTTAACTGCTGGGCTGTGTAAACGAGCACTAAGGAGCCAA | 8820 |
| Db | 8761 | ACATGAGGAGGAGGGGTATGTTTTAACTGCTGGGCTGTGTAAACGAGCACTAAGGAGCCAA | 8820 |
| QY | 8821 | TGTAGACATTTTGACTTAAGAAAGATCATATCAAAAGCCGGGGGAGGAGGTATGAGAGTTGG | 8880 |
| Db | 8821 | TGTAGACATTTTGACTTAAGAAAGATCATATCAAAAGCCGGGGGAGGAGGTATGAGAGTTGG | 8880 |
| QY | 8881 | ACTACAGTGTCTCAAGACCCCCATAGAGAGCCAGTTTCCCTTCTTCCCTGAGGCTCAAGC | 8940 |
| Db | 8881 | ACTACAGTGTCTCAAGACCCCCATAGAGAGCCAGTTTCCCTTCTTCCCTGAGGCTCAAGC | 8940 |
| QY | 8941 | CTGGCTGCAGCGCCACGTGCTCTACATAGCCTTCTCTCTAGGCTGTGCACCATG | 8995 |
| Db | 8941 | CTGGCTGCAGCGCCACGTGCTCTCAATAGCCTTCTCTCTAGGCTGTGCACCATG | 8995 |

| | | | | |
|------------|---|-----------|-----|------------------------|
| | RESULT 2 | | | |
| | AC098570 | | | |
| LOCUS | AC098570 | 192060 bp | DNA | linear HTG 20-AUG-2002 |
| DEFINITION | Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered pieces. | | | |
| ACCESSION | AC098570 | | | |
| VERSION | AC098570.2 GI:22325297 | | | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. | | | |
| SOURCE | house mouse. | | | |
| ORGANISM | Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriongnathi; Muridae; Murinae; Mus. | | | |
| REFERENCE | 1 (bases 1 to 192060) | | | |
| AUTHORS | Birten,B., Nusbaum,C. and Lander,E. | | | |
| TITLE | Mus musculus, clone RP23-278N11 | | | |
| JOURNAL | Unpublished | | | |
| REFERENCE | 2 (bases 1 to 192060) | | | |
| AUTHORS | Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barin,N., Bastien,V., Boguslavsky,L., Boukhalgalter,B., Brown,A., Camarata,J., Campodiano,A., Chang,Y.J., Chazarro,B., Choepel'V., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., DeKrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Grinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lepoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,H., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menues,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhhang,P., Pierre,N., Pollara,V., Raymond,C., Reta,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiljev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.-J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M. | | | |
| TITLE | Direct Submission | | | |

JOURNAL

Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192060)

REFERENCE

Bliren,B., Nushbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Bairn,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Chopel,Y., Collymore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karakas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Milhova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Notbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuppback,R., Seaman,S., Severij,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testife,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 20, 2002 this sequence version replaced g1:16356898.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MTBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L14796

Center clone name: 278.N.11

----- Summary Statistics -----

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 190464 bases at least Q40

Consensus quality: 191061 bases at least Q30

Consensus quality: 191345 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 191560; sum-of-contigs

Quality coverage: 12.4 in Q20 bases;

Quality coverage: 10.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 4921: contig of 4921 bp in length

* 4922 5021: gap of 100 bp

* 5022 15200: contig of 10179 bp in length

* 15201 15300: gap of 100 bp

* 15301 78692: contig of 63392 bp in length

* 78693 78792: gap of 100 bp

* 78793 116118: contig of 37326 bp in length

* 116119 116218: gap of 100 bp

* 116219 174491: contig of 58273 bp in length

* 174492 174591: gap of 100 bp

* 174592 192060: contig of 17469 bp in length.

Location/Qualifiers

1. 192060

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="RP23-278N11"

/clone_lib="RPCI-23 Female Mouse BAC"

FEATURES

source

[illegible]

| | | | |
|----|-------|---|-------|
| OY | 1861 | CAGATGGCTCAAGCTTCCACAGGAGCACTTGCTGCTCTTGCAGAGAGGACCTAGATTCAAGTCCC | 1920 |
| Db | 42021 | GAGATGGGCTCAGGCTTCCAGAGGACACTTGCCTGCTCTTGCAGAGGAGGACCTAGATTCAAGTCCC | 42080 |
| OY | 1921 | AGGACTCGATTGSGTGGCTCACAGCCCATCGTAAATCCAGTTTCCAGAGGTTCCACACCTT | 1980 |
| Db | 42081 | AGGACTCGATTGSGTGGCTCACAGCCCATCGTAAATCCAGTTTCCAGAGGTTCCACACCTT | 42140 |
| OY | 1981 | CTTCTGGCCCTCCACAGGAGCCACATACATAGTATGACACAGACATACATATGACAGCAAAACAC | 2040 |
| Db | 42141 | CTTCTGGCCCTCCACAGGAGCCACATACATAGTATGACACAGACATACATATGACAGCAAAACAC | 42200 |
| OY | 2041 | CATTACACATTAATAATAATAAGGAACTTAAAGGTGCATGTGTGTAATAACATTTGTGCT | 2100 |
| Db | 42201 | CATTACACATTAATAATAATAAGGAACTTAAAGGTGCATGTGTGTAATAACATTTGTGCT | 42260 |
| OY | 2101 | TACACATGCTGATTTAAGACATGTTCAAACGCACACACTGAGAGAGGATCTGGGGCTGGAG | 2160 |
| Db | 42261 | TACACATGCTGATTTAAGACATGTTCAAACGCACACACTGAGAGAGGATCTGGGGCTGGAG | 42320 |
| OY | 2161 | AGATGGCTCAGGGGTTAAGAGACACTGACGTGCTTCCGGAAGGAAGTCCCTGAGTTCAAT | 2220 |
| Db | 42321 | AGATGGCTCAGGGGTTAAGAGACACTGACGTGCTTCCGGAAGGAAGTCCCTGAGTTCAAT | 42380 |
| OY | 2221 | CCTACCAACACACATGSGTGGCTCACCAACCATCCATTAATGAGATCTGACACCCCTCTGCTGT | 2280 |
| Db | 42381 | CCTACCAACACACATGSGTGGCTCACCAACCATCCATTAATGAGATCTGACACCCCTCTGCTGT | 42440 |
| OY | 2281 | GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTTAAATATCTAATAATAATCTTTT | 2340 |
| Db | 42441 | GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTTAAATATCTAATAATAATCTTTT | 42500 |
| OY | 2341 | TTTAAAAAATGAAGGAGTCTGAGACACCCCAAAAGAGATTATGACAGAGTCACTCACG | 2400 |
| Db | 42501 | TTTAAAAAATGAAGGAGTCTGAGACACCCCAAAAGAGATTATGACAGAGTCACTCACG | 42560 |
| OY | 2401 | GGTGATTATCTTCTGAGGATTTTCTTCTTCCGCTTGGCTTGGCACTGSGTGGACAGACG | 2460 |
| Db | 42561 | GGTGATTATCTTCTGAGGATTTTCTTCTTCCGCTTGGCTTGGCTTGGCACTGSGTGGACAGACG | 42620 |
| OY | 2461 | CCCCCTTTTCATTACAGAAGCGGGTGTACATTTATTTCTGAACAACAGCACCTGCGAT | 2520 |
| Db | 42621 | CCCCCTTTTCATTACAGAAGCGGGTGTACATTTATTTCTGAACAACAGCACCTGCGAT | 42680 |
| OY | 2521 | ATGTTTACTGTCCTTGTCTGACTATGACAGCGGGGCGCGGGG - -CACACACACACAC | 2578 |
| Db | 42681 | ATGTTTACTGTCCTTGTCTGACTATGACAGCGGGGCGCGGGGCGCGCACACACACAC | 42740 |
| OY | 2579 | ACACACACACACACACACACACACACACATTCACGTCTTCACAGGCTCTTGGGAAAGT | 2638 |
| Db | 42741 | ACACACACACACACACACACACACACACATTCACGTCTTCACAGGCTCTTGGGAAAGT | 42800 |
| OY | 2639 | CAAGAAGAGGCTGCCCTCAACACAGATCTTCATCTTTCCTTCCTAAGGAGACACAGATT | 2698 |
| Db | 42801 | CAAGAAGAGGCTGCCCTCAACACAGATCTTCATCTTTCCTTCCTAAGGAGACACAGATT | 42860 |
| OY | 2699 | CCAAGGTGGCAGAAATCTACAGGGGGGCGAGAGGCGAGGAGGGGGAGACAGCGCATAGTTT | 2758 |
| Db | 42861 | CCAAGGTGGCAGAAATCTACAGGGGGGCGAGAGGCGAGGAGGGGGAGACAGCGCATAGTTT | 42920 |
| OY | 2759 | CCAGAGACCTTACAGCAGAGGCGCAGCAAGGAGATCCCAAGCTCCAGGGCAGGGCAGGAGTGG | 2818 |
| Db | 42921 | CCAGAGACCTTACAGCAGAGGCGCAGCAAGGAGATCCCAAGCTCCAGGGCAGGGCAGGAGTGG | 42980 |
| OY | 2819 | GGCCCTTGTTCGAGAGAGGAGGCGGAGAACAGGCTTCAAAGGCAACAGGTTTATGG | 2878 |
| Db | 42981 | GGCCCTTGTTCGAGAGAGGAGGCGGAGAACAGGCTTCAAAGGCAACAGGTTTATGG | 43040 |
| OY | 2879 | CAGCTCATTAAGAAGTGGAGGTGCTGGCTCACTTAGAAGAAGGAGAAAGAAGGCAAGCCCTT | 2938 |
| Db | 43041 | CAGCTCATTAAGAAGTGGAGGTGCTGGCTCACTTAGAAGAAGGAGAAAGAAGGCAAGCCCTT | 43100 |

| | | | |
|----|-------|--|-------|
| OY | 2939 | GTCCCACTGAGCGGAGGTCATGCTGAGTAGAGAGATCTGAGAGGGTGCAGAGAGCC | 2998 |
| Db | 43101 | GTGGCCACTGAGCGAGGGTCATGCTGAGTAGAGAGATCTGAGAGGGTGCAGAGAGCC | 43160 |
| OY | 2999 | ACCTGCTGTGCCAAGAGAAACCCCAAGTGTAACTCTGGCCCTTGGGTGCTGAATTCCAGC | 3058 |
| Db | 43161 | ACCTGCTGTGTCCCAAGAGAAACCCCAAGTGTAACTCTGGCCCTTGGGTGCTGAAGTTCCAGC | 43220 |
| OY | 3059 | TACAAGACCCCAAGAGAGTCTTACTCCATCCCATTCACAGTGCCTCCCTGCGCCGACACCC | 3118 |
| Db | 43221 | TACAAGACCCCAAGAGAGTCTTACTCCATCCCATTCACAGTGCCTCCCTGCGCCGACACCC | 43280 |
| OY | 3119 | CACCCCGCACTCCGTGCCACTTCTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGGTT | 3178 |
| Db | 43281 | CACCCCGCACTCCGTGCCACTTCTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGGTT | 43340 |
| OY | 3179 | GCCCTACGTCAGAGGTAGAGCCCAAGGTCCTAAGCCGGAAGTGCACCCATCCCTGAACCTGCA | 3238 |
| Db | 43341 | GCCCTACGTCAGAGGTAGAGCCCAAGGTCCTAAGCCGGAAGTGCACCCATCCCTGAACCTGCA | 43400 |
| OY | 3239 | GAGCCAAAGGGGGGCGACACAGCGACACTCAGAGCTGTACAGCTTGTGCTGGGCTCTAGATTTC | 3298 |
| Db | 43401 | GAGCCAAAGGGGGGCGACACAGCGACACTCAGAGCTGTGTACAGCTTGTGCTGGGCTCTAGATTTC | 43460 |
| OY | 3299 | CCAGGACCTGGGCACTACTTCCCACACCCCCCATTCACATTCCTCTGGGGCCCTATCTT | 3358 |
| Db | 43461 | CCAGGACCTGGGCACTACTTCCCACACCCCCCATTCACATTCCTCTGGGGCCCTATCTT | 43520 |
| OY | 3359 | CCCTTATATGTGTGAAGAAATTCCCTGGGGGGGGGGGGTGTGTGTATAGACAAAGTCTCT | 3418 |
| Db | 43521 | CCCTTATATGTGTGAAGAAATTCCCT - GGGGGGGGGGGTGTGTGTATAGACAAAGTCTCT | 43578 |
| OY | 3419 | CGGTCTCCTGCAGCGAGCTGGCCACACACTCTCTAAGATCTCCGAGGTGGTGCCTCCTCT | 3478 |
| Db | 43579 | CGGTCTCCTGCAGCGAGCTGGCCACACACTCTCTAAGATCTCCGAGGTGGTGCCTCCTCT | 43638 |
| OY | 3479 | TCCACACAGGTAAAGGCAATTGGTGGGGGACACATGCTGACCAAGTGGTGGAGGGAC | 3538 |
| Db | 43639 | TCCACACAGGTAAAGGCAATTGGTGGGGGACACATGCTGACCAAGTGGTGGAGGGAC | 43698 |
| OY | 3539 | AGGGTCTTGTCTCTCTGTGGAGGCTGTGCTTCTGTAGCACCTTGGTATAAATTTGGG | 3598 |
| Db | 43699 | AGGGTCTTGTCTCTCTGTGGAGGCTGTGCTTCTGTAGCACCTTGGTATAAATTTGGG | 43758 |
| OY | 3599 | GGTGAAGTAAAGTGCCTGCGAAACCTCGAAAGAAGCAAGAAGCACAGGCTCTTGGGC | 3658 |
| Db | 43759 | GGTGAAGTAAAGTGCCTGCGAAACCTCGAAAGAAGCAAGAAGCACAGGCTCTTGGGC | 43818 |
| OY | 3659 | CTTCAATGAAGAACTTACACAGACCCCTTCTCTGTAAAGTCACTTCCGCTCATCTGTGT | 3718 |
| Db | 43819 | CTTCAATGAAGAACTTACACAGACCCCTTCTCTGTAAAGTCACTTCCGCTCATCTGTGT | 43878 |
| OY | 3719 | AGATTTCCTGGGACCAAGTGGCTCCTGGGACTCAGATTTCTACAAATTAAATTCAGGACA | 3778 |
| Db | 43879 | AGATTTCCTGGGACCAAGTGGCTCCTGGGACTCAGATTTCTACAAATTAAATTCAGGACA | 43938 |
| OY | 3779 | GTCTGAGACTTGGAGCTCCGCTCGCTGTATTATTAATCTACTTCTCTGGCTGCATTTCTGT | 3838 |
| Db | 43939 | GTCTGAGACTTGGAGCTCCGCTCGCTGTATTATTAATCTACTTCTCTGGCTGCATTTCTGT | 43998 |
| OY | 3839 | GTTCATGTCTTACACATCTGAATAGTTCTTGTGTGCACCAATTCCTCCGACACTCTGGG | 3898 |
| Db | 43999 | GTTCATGTCTTACACATCTGAATAGTTCTTGTGTGCACCAATTCCTCCGACACTCTGGG | 44058 |
| OY | 3899 | GAGGTCTGATCTTGTGGCACATGTATTCCTGTGGATGTAAAGTGTGAGCCACACAGAGAGGG | 3958 |
| Db | 44059 | GAGGTCTGATCTTGTGGCACATGTATTCCTGTGGAGTGTAAAGTGTGAGCCACACAGAGAGGG | 44118 |
| OY | 3959 | GGAGGTGAGAGAGCTGTGTGCCCTTAAGGCCATTAAGGCTGTGACATCAACCCCTTCTAGAA | 4018 |
| Db | 44119 | GGAGGTGAGAGAGCTGTGTGCCCTTAAGGCCATTAAGGCTGTGACATCAACCCCTTCTAGAA | 44178 |
| OY | 4019 | ATGGCCCTCCATTTTTTGGTGTACCATGATCTAATTTTATACAGATGGGCGAGTAAMAC | 4078 |

[illegible]

| | | | |
|----|-------|---|-------|
| QY | 6170 | AGAGGGCTGAGCACCAACGATGACAAGTTCCTGTGTCTCATGCGACGAGATTCCTGGCC | 6229 |
| Db | 46391 | AGAGGGCTGAGCACCAACGATGACAAGTTCCTGTGTCTCATGCGACGAGATTCCTGGCC | 46450 |
| QY | 6230 | AGTTTCAAGGACATGAGACTCATCTGGGGGAAACAAGTATGCCAAGGCTTAAGGCC | 6289 |
| Db | 46451 | AGTTTCAAGGACATGAGACTCATCTGGGGGAAACAAGTATGCCAAGGCTTAAGGCC | 46510 |
| QY | 6290 | CATTTTGGTCTTAATTAATCAAGAACCCCTGGGGATGACAGGCTCTGAGCAGACGACTTT | 6349 |
| Db | 46511 | CATTTTGGTCTAATTAATCAAAACCCCTGGGGATGACAGGCTCTGAGCAGACGACTTT | 46570 |
| QY | 6350 | TTAAAAAGCTCCAGAGTATTCGTATGACAGACCTGGGAACAACACAGCTACAGGTTCAAA | 6409 |
| Db | 46571 | TTAAAAAGCTCCAGAGTATTCGTATGACAGACCTGGGAACAACACAGCTACAGGTTCAAA | 46630 |
| QY | 6410 | CAGAAAGAGCAAACTGAGGAAAGCTTGGGATGGGGAGCCTTCTTCCAGGCCAGTAAAT | 6469 |
| Db | 46631 | CAGAAAGAGCAAACTGAGGAAAGCTTGGGATGGGGAGCCTTCTTCCAGGCCAGTAAAT | 46690 |
| QY | 6470 | GGAGGCTGGTTTGGACAGTGGTGGACACTTCTCTCCGTGATATAGCTATATCCATCCACT | 6529 |
| Db | 46691 | GGAGGCTGGTTTGGACAGTGGTGGACACTTCTCTCCGTGATATAGCTATATCCATCCACT | 46750 |
| QY | 6530 | CATCATCCATACACCCACCCATCCATTTATGACACCATCTTCCATCATCATCATATC | 6589 |
| Db | 46751 | CATCATCCATACACCCACCCATCCATTTATGACACCCATCTTCCATCATCATCATATC | 46810 |
| QY | 6590 | CAGCTAACCCACCCAGGACATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTTC | 6649 |
| Db | 46811 | CAGCTAACCCACCCAGGACATCCATCCAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTTC | 46870 |
| QY | 6650 | CTTCACTCATTTTATATTCACAAGAAAGCTGGTATTTACTAAATGTGGGAGATTTAAT | 6709 |
| Db | 46871 | CTTCACTCATTTTATATTCACAAGAAAGCTGGTATTTACTAAATGTGGGAGATTTAAT | 46930 |
| QY | 6710 | TAAATTTTGAAGAGCTCTGTTGATTTGATGCTGATTTGATGAGACAGGTATACCA | 6769 |
| Db | 46931 | TAAATTTTGAAGAGCTCTGTTGATTTGATGATTTGATGATGATGAGACAGGTATACCA | 46990 |
| QY | 6770 | CAGCACACGCTGTGGCAATTCGGGAAAGTTTGGGATGTTTCTCTCTTCCACCGTGTG | 6829 |
| Db | 46991 | CAGCACACGCTGTGGCAATTCGGGAAAGTTTGGGATGTTTCTCTCTTCCACCGTGTG | 47050 |
| QY | 6830 | GGTTTGGGGATTTGAACCTCAATTTTCGGGGCTGGTGGCAAGCTCTTTTACCAACGACCA | 6889 |
| Db | 47051 | GGTTTGGGGATTTGAACCTCAATTTTCGGGGCTGGTGGCAAGCTCTTTTACCAACGACCA | 47110 |
| QY | 6890 | TTTTCGTCGACATCATTTATTTAGAAAGCATCTTATGTAGTCCAGGCTGGGCTCAAGC | 6949 |
| Db | 47111 | TTTTCGTCGACATCATTTATTTAGAAAGCATCTTATGTAGTCCAGGCTGGGCTCAAGC | 47170 |
| QY | 6950 | TTTTCATGTGCGCACGAGATGACCTTAACTCTGCTCTTCCAGGCTCCACCGGAGTGTCTA | 7009 |
| Db | 47171 | TTTTCATGTGCGCACGAGATGACCTTAACTCTGCTCTTCCAGGCTCCACCGGAGTGTCTA | 47230 |
| QY | 7010 | GGTTTACAGGTTTCAAC----- | 7027 |
| Db | 47231 | GGTTTACAGGTTTCAAC----- | 47290 |
| QY | 7028 | -----TGCTGAATGCCCTTTAATCCAGCACTGTGGG | 7060 |
| Db | 47291 | AGAAATCTGGCAACAGCTGGGCTGAGTGAGTGCTGAATGCCCTTTAATCCAGCACTGTGGG | 47350 |
| QY | 7061 | GGGGGGGGGGAGCGGATCCTGAGCTTGGAGGCCAGTTTGGTCTACAGAGTTTACAGAT | 7120 |
| Db | 47351 | GGGGGGGGGGAGCGGATCCTGAGCTTGGAGGCCAGTTTGGTCTACAGAGTTTACAGAT | 47410 |
| QY | 7121 | ACCTGGGGCTATACAGGGAAACCTTATCCCAAAACAACAACAACAACAACAATAATATT | 7180 |
| Db | 47411 | ACCTGGGGCTATACAGGGAAACCTTATCCCAAAACAACAACAACAACAACAATAATATT | 47470 |

[illegible]


```

|||||
Db 48551 TATGAGATATTTTGTGGAGTTTTCACCTTCCCTGTGGGCTGCCATTAACCTCAG 48610
QY 8313 CTCCTGGGGCTAGTAGCAATGCCCTTCACTGATGAGCCATCTCGCTCCCTGCTGCCA 8372
Db 48611 CTCCTGGGGCTAGTAGCAATGCCCTTCACTGATGAGCCATCTCGCTCCCTGCTGCCA 48670
QY 8373 CCTCCTCTTTATTTTCCAGATGGGACTAGCAGCTGACCTGAGGCTTAAGCTCACCAGTCA 8432
Db 48671 CCTCCTCTTTATTTTCCAGATGGGACTAGCAGCTGACCTGAGGCTTAAGCTCACCAGTCA 48730
QY 8433 TCCAGAGTGGCTAGCCAGGAGGAGCTCAGGAGTATGCTGCGCTGCTCCACAGTGTAG 8492
Db 48731 TCCAGAGTGGCTAGCCAGGAGGAGCTCAGGAGTATGCTGCGCTGCTCCACAGTGTAG 48790
QY 8493 AATTACAGGCTACATCACTGCTGGAACATTTTAACTGAACTCTGAGATAGACGAG 8552
Db 48791 AATTACAGGCTACATCACTGCTGGAACATTTTAACTGAACTCTGAGATAGACGAG 48850
QY 8553 CACTCTACCAATGAGAGGCTCTTTTGTGTGGTTGGTTGCTTCCCTGCTCATAGATCAGG 8612
Db 48851 CACTCTACCAATGAGAGGCTCTTTTGTGTGGTTGGTTGCTTCCCTGCTCATAGATCAGG 48910
QY 8613 CAGTCTGAATAGTGTAGCTTGGGCTACATACATCTTGTCTCAAAAAGCCTATAGAGT 8672
Db 48911 CAGTCTGAATAGTGTAGCTTGGGCTACATACATCTTGTCTCAAAAAGCCTATAGAGT 48970
QY 8673 AGGAGGCTGAGGCTTAAGAAGAGCCTTAAGCCGCTGTGATAGCACAGAGATAGCCTG 8732
Db 48971 AGGAGGCTGAGGCTTAAGAAGAGCCTTAAGCCGCTGTGATAGCACAGAGATAGCCTG 49030
QY 8733 CACTATATAGAGAGACCTGTTTCAAAAACATGGAGGAGGCTATGTTTAAAGCTGG 8792
Db 49031 CACTATATAGAGAGACCTGTTTCAAAAACATGGAGGAGGCTATGTTTAAAGCTGG 49090
QY 8793 GCTGTGTACAGGAGCTAAGGAGGAGCAATGTAGACATTTGACTAAGAAGATCATCATC 8852
Db 49091 GCTGTGTACAGGAGCTAAGGAGGAGCAATGTAGACATTTGACTAAGAAGATCATCATC 49150
QY 8853 AAAGCCGGGTGGGAGGCTAAGGTTGAGCTACAGTGTCTCAAGACCCCATAGAGGCCA 8912
Db 49151 AAAGCCGGGTGGGAGGCTAAGGTTGAGCTACAGTGTCTCAAGACCCCATAGAGGCCA 49210
QY 8913 GTTTCCTCTTCTCTCTGGGCTCAAGCTGCTGAGGCGACAGCTCTACATGCTCTT 8972
Db 49211 GTTTCCTCTTCTCTCTGGGCTCAAGCTGCTGAGGCGACAGCTCTACATGCTCTT 49270
QY 8973 CTCTCTAGGCTGCTCCACCATG 8995
Db 49271 CTCTCTAGGCTGCTCCACCATG 49293

RESULT 3
AC101018 149109 bp DNA linear HTG 21-AUG-2002
LOCUS Rattus norvegicus clone RP32-328P7, *** SEQUENCING IN PROGRESS ***,
DEFINITION 9 unordered pieces.
ACCESSION AC101018
VERSION AC101018.2 GI:22381320
KEYWORDS HTG; HTGS_PHASE1; HTGS_FUZZTOP; HTGS_ACTIVEFIN.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 149109)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Rattus norvegicus, clone RP32-328P7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149109)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Birn,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardina,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menes,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliou,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17059792.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14794
Center clone name: 328_P_7

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 46829: contig of 46829 bp in length
* 46830 46929: gap of 100 bp
* 46930 49621: contig of 2692 bp in length
* 49622 49721: gap of 100 bp
* 49722 70009: contig of 2028 bp in length
* 70010 70109: gap of 100 bp
* 70110 77196: contig of 7087 bp in length
* 77197 77296: gap of 100 bp

```


Db 100265 AGATTTTACAAATTAATCTCAGACAGACGTCCTGAGAAAGCTGGAAATTAAGTGTCTCCT 100324
Qy 3799 TGCCGTATTACTACTTCTCTGCGTCGTCATTTCTGTGATCTTACATCTG 3858
Db 100325 TGGCAACATTTCTTCTTCTCTCTGAGGTTCAATTTCTTGTCTATCTTGTGACCTTCTG 100384
Qy 3859 AAATGC-----TTTCTTGTGTCAACATTTCCCTGACACTCCCTGGAGCGTATCTTGG 3914
Db 100385 AAATGTTTCTTCTTGTGTGTGACACATTTCCCTGACACTCTGTGAGAGTTGGGCTTGG 100444
Qy 3915 CAC-----ATGATCTCTGGAGATTAAGTGCACACCACAGCA-GAGAGGGGAGAGTACG 3969
Db 100445 CCCACCATGATTCCTGGAGATGTAAGCTGCATCCACAGAGGGGAAAGGCAAGCTGG 100504
Qy 3970 AG-----CTGTCTCTAGAGCCCTATTAGCCCGGACATCACCCCTTCTCTGAATAGGCC 4025
Db 100505 AGGCTACAGGCTCTAGGCCCTCTCTGCGCTGTGATCATCTCTGCGCTTGAATAGGCC 100564
Qy 4026 CTCATTTTTCGGTTTACCATGATCTATTTATATCAGAGTGGGAGTGAAGCCAAACT 4085
Db 100565 CTGTCTCTCAGTCTCCGTTGTCTATTT--TATCAGAGCTGACAGTAAGAAAGCAAGCT 100622
Qy 4086 GCCCAGAAAGTTGGGACTCACTCAGACCAAGGTTATCTGCTCAGAAATCCCGCTGTACT 4145
Db 100623 GCCCAGAAATTTTGGGACTCACTTAACCAAGGTCATATGCTCAGAGCTCTTTTGTCACT 100682
Qy 4146 T-GAGTTGGAGAAATCGCTCTGGGGGCTCCAGGCTTGGTTTACAGAGGATTC 4204
Db 100683 TGGAGTTGGAGATATCTCTCTCCAGGGGCTCCAGGCTCTGTTTACAGAGAGGCAATCC 100742
Qy 4205 TTTGTATAG---GCATGACCTAGTCTATGTGTGTACTACATTTGCTCTCAGTTAAAGC 4261
Db 100743 TTTGTATAGGAACCTGTACCTAACCATGTGTGTACTACACATCTCTGCTCAGTTAAAGC 100802
Qy 4262 TGGAACTAAACCACGCGCAGCGCCAGAGTTCTCTACAGTTGTACCCCAAGAACAA 4321
Db 100803 TGGGACATAAACCCGTCGACACCCAGGATTCCTTATGTGTGACCTCCACAGAACAG 100862
Qy 4322 GACAGTATATGCAAGATAGTATAGTGGGAGAGAAAGAACTTAA-ACCCCCCAAG 4380
Db 100863 GACAGTATATGCAAGAGAGACAGTATGAGAGAGAGATCTTAAAGTCTTAAGGCCAT 100922
Qy 4381 GCCCAGAGTTCCGTTCCCTAGTTACAAATGCCAGTATGAGTGTAG-----CTACT 4432
Db 100923 AGTCTCTCTGCTCCGTTCCCTAGTTCTGAATGTGATGTGATGCCAATTTGATTTACT 100982
Qy 4433 ATGGGCTGTGAGTTGTGATGATCAGACATGATGTA-ITGTCATGTGTGTGTGTATA 4491
Db 100983 ATGGGCTGTGAACTGTGATGATCAGAGTGTGATGATGATGATGTGTGTATA 101042
Qy 4492 TCTGAG-CACCTTGGAGGCTGAAGCAGGAGATTGCTATATGTTGAGGCGAGCTGAGC 4550
Db 101043 TCTGAGCCTGAGAGGCGGAGCATG-----TTTCTATGTTTGGAGCCACTGAGC 101097
Qy 4551 TATGAGCGAGACTTGTCTTTTAAAGAAATAATGAAGCCAGCAGTGTGTGACAGGCC 4610
Db 101098 TATGAGGAGAGAGCTGTCTTTAAAGAAATAATGAAGCAACACG----- 101144
Qy 4611 TTTTATCCAGCACTTGGAGGAGAGAAAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGG 4670
Db 101145 ----- 101144
Qy 4671 TCTATAGATGATTCAGAGCAGCAGCGGCTACACAGAAACCTGTTTGAAAAAC 4730
Db 101145 ----- 101144
Qy 4731 AGAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 4790
Db 101145 -----ACCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 101189
Qy 4791 CTCTATCTCTAGGCTGTGTCTGTCTGTAGTGTGATGAGTTTGGGAGCTTCAAGCTTATATA 4850
Db 101190 TTTCTATCTGTAGGTTGTGTCTGTCTGTAGTGTGAGCTTGGGCTTCAAGACTTATATA 101249

Qy 4851 TAAATAGCCCTTTTATCAGTGTACAGACAGAAAGTTTCACTGTGAGACAGTGG 4910
Db 101250 TAGGCCCTTTTATCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 101309
Qy 4911 GACCCGTGAAGAACTACTCTTGTCCAGCCCAAAAAATCTGTGGAAGCTTCTGTGAGAAAT 4970
Db 101310 GTCCCTGAGAAAGACTCTTGCACAGCTTGTGAAATTCGTGGAAGGCTTCTGTGAGAAAT 101369
Qy 4971 GTGTCT-----CCCATCA 4982
Db 101370 GTCCCGATTGATGTGTGAGAAATAATGAAAGTGGGAAAGCCAAAGACAGACAGCTGTAG 101429
Qy 4983 GACTACTGTCTGAAGAGCAGAGAGAGGTTGGAAGATGT-----T 5025
Db 101430 AACTACTGTCTGAAGAGCAGAGAGAGGCTTGAAGAGTGTCTGTACGAGACAGAGCAT 101489
Qy 5026 GTGTGACAGACAGTTGGAACAGAGACAG---AGGGGAGGCATTCAGATTTCTGAA 5081
Db 101490 GTGTGACAGACAGTTGGAACAGAGAGGACTTGGGACAGGGGAAACCGTCAAGATTTCTGA 101549
Qy 5082 CATGTACTGACTTTTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5141
Db 101550 CACATAGCTGT 101597
Qy 5142 AGGGGACAGGGGTGAGACCAATGATTCAGTTGAGGGACATCCAGCCAGGGTCTT 5201
Db 101598 TGGGAGACAGGGGTGAGATCAGT 101657
Qy 5202 GTGTGACAGCTTAAAGATGAGAGCCCTTACCCCTCTGAGTTTGGGAGACAGAG 5261
Db 101658 GTGTGACAGCTTAAAGATGAGAG--CTTCAACACCTCTTGAAGGAGACAGTGGG 101716
Qy 5262 AGCTGAGAGATCTCTGT 5320
Db 101717 AGCTGAGAGATCTCTGT 101776
Qy 5321 AGCAGAGCA-----GTTGACAGTACCCTTCAAGACAGGACATC 5362
Db 101777 AGCAGAGCACTTGT 101836
Qy 5363 CCCTTGT 5422
Db 101837 CCACCTTGT 101895
Qy 5423 GAGAGGACAGAGCTTGT 5482
Db 101896 GAGAGGACAGAGCTTGT 101954
Qy 5483 TAGTTTATTTTACTGT 5542
Db 101955 TAGTTTATTTTGT 102014
Qy 5543 CTCACCAAGTATGCCCCCATCTTCCACAGAGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAG 5602
Db 102015 CTCACCAAGTATGCCCCCATCTTCCACAG--AGTGTCAAAAGAGAGAGAGAGAGAGAGAG 102073
Qy 5603 AGGTTAACAAGTATGAG 5662
Db 102074 AGAGTAAACAAGTATGAG 102133
Qy 5663 TTTGTTTGTGTGAACCGAG 5722
Db 102134 -----TTGCTGAACAGGA----- 102147
Qy 5723 GT 5782
Db 102148 -----GTGAGT 102189
Qy 5783 AGAATATCTATATTTGGGCGCCATGCCAAGAGGATTTTGGGAGAGCT-----AG 5833
Db 102190 GGAATATCTATATTTGGGCGCCATGCCAAGAGGATTTTGGGAGAGAGAGAGAGAGAGAGAGAG 102249

| | | | |
|----|--------|---|--------|
| OY | 5834 | GCCTCGCAACACAGTAAGCTGCCCAAGATGATTTGGTGCCCTGAATCCACAAAGGGCA | 5893 |
| Db | 102250 | GCCTTCGCAACACACAGCAAGCTGCCCAAGATGATTTGGTGCCCTGAATCCACAAAGGGCA | 102309 |
| OY | 5894 | GGCTCATGAGGTGGACGAACATCAACAAGTAAGCCACCCCTGGGGGCTCACAAAGGG | 5953 |
| Db | 102310 | CGCTATCAGAGTGGACGAAG--CACAGATTAAGCCACCCCTGGAGGGCTCAGA-----G | 102361 |
| OY | 5954 | AGTTTCAAGAGGTAAAGGCCCAAGCCATTTATATCCAAAGCATGATCAAAATCAAAAGT | 6013 |
| Db | 102352 | AGTTTCAAGAGGTAAAGGCCCAAGCCATTTATATCCAAAGCATGATCAAAATCAAAAGT | 102421 |
| OY | 6014 | GCACGAGAGATTTAGCTGGAGAGATGGGGGCTGTACGTGGGGACACTGACCTTGCACCTT | 6073 |
| Db | 102422 | GCACGAGAGATTTAGCTGGAGAGATGGGGGCTGTCAATGTGGGATCTGACTTGGCCCTT | 102481 |
| OY | 6074 | ATTACTCATAGGCCCAAGAGACATCAACAAGGGGTGACTGGGTCCTACTAGCTGGAGC | 6133 |
| Db | 102482 | GTTTACTCTAGGCCCAAGAGACATCAACAAGGGGTGACTGCTCTCATAGCTGGGGG | 102541 |
| OY | 6134 | AGGCACGTGAGAAATGGGTGACCTTCATCCGATGAGAGAGGGCTGAGACACACAGGTAC | 6193 |
| Db | 102542 | AGGCACATAGAAATGGGTG-----TGAGAGGGGTTGAGAAATGACCAGGTGC | 102589 |
| OY | 6194 | AAAGTTTCCC---TGCTCTCATGCCAGATTCCTGGCCAGTTTTCAAAGAGCTAAGGAC | 6250 |
| Db | 102590 | AAAGTTTCTCTGATGGTGCCCTATCTGGGATTTCCACACAGTTTTCAAAGAGCCAGTGC | 102649 |
| OY | 6251 | TCATCTCGGTGGGAAACAAAGTATTCCAAGCCCTAAGCCCATTTTGGTCTAATTAATCA | 6310 |
| Db | 102650 | TCATCTCGATGGAATTAAGATATCCAAAGCCCTAAGCCCTATTCGTCTAATTAATCA | 102709 |
| OY | 6311 | GAACCCCTGGGGGATGACAGCTCTGACAGCAGCAGAGCTTTTAAAAAAGCTCCACAGGTAT | 6370 |
| Db | 102710 | GAACCTCTGGGGATGACAGCTCTGACAGCAGG----TTTTTAAAGAGCTCCACAGGCACT | 102755 |
| OY | 6371 | CTGATCAGC--AGCTGGAACAACACAGCTACAG--TTCAACAGAAAGAGCCAAAGCT | 6426 |
| Db | 102766 | CTGATCAGCCGAATTCAGATTAATAAACAGCTACAGGGTTCAACAGAAAGAGCCAAAGCT | 102825 |
| OY | 6427 | AGGAAACCTTGGGATGGGAGACCTCTCTCCAGGCCAGTAGATGAGAGCTGGTTAGCAGT | 6486 |
| Db | 102826 | AGGAAACCTTGGGAT--GGAAGCTTCTTTTCAGCCCGTAGACCAAGGCCACCAAAANN | 102883 |
| OY | 6487 | GGTGCACACTCTCTCTCTCTCATATAGCTATTCATCCATCATCATCATACACC | 6546 |
| Db | 102884 | NN | 102943 |
| OY | 6547 | ACCATTCATTATGACCCATCCCTTCATCATCATCATCATCATCCACCCACG | 6606 |
| Db | 102944 | NN | 102988 |
| OY | 6607 | ATCCATCCAAACCTTCCCTTCTCTCTCTCTTCTTCTTTTCCCTTACATTCATTTA | 6666 |
| Db | 102989 | ACCATTCCAAACTTGGCTTCTCTTCTCTTCACTTCACTCTCTTCCCTTCACTCATTC | 103048 |
| OY | 6667 | TCCAACACAGAACTGGTATGTACTAATGAGGAGATTAATTAATTTTACAGCTC | 6726 |
| Db | 103049 | CGTAACCAAGAACTGTTTACTAATAATGAGAGATTAATTAATTAATTAATTA | 103106 |
| OY | 6727 | GTTGATGACTGATTTGTCATGATGTCGACAGGTATACATACACAGCACAGTGTGGCAA | 6786 |
| Db | 103107 | ATTATATGACTGATTTGTCATGTCGTGGGATATACATGCGACAGTACATGTAAGAA | 103166 |
| OY | 6787 | TCGAGAAAGGTTTTGGGTGTTGTTTCTCTCTCCACCGTGGGTTTC--TGGGATTTGAA | 6845 |
| Db | 103167 | TCAGAGAAAGTTTTGGGTGTTGTTTCTCTTCCACTGTGTGGGTTCTTCGAGACTTAC | 103226 |
| OY | 6846 | CTCAAAATATGCGGGCTGGCAAGTGTCTTA--CAACGAGCCATTTTCTCAGACATC | 6904 |
| Db | 103227 | CTCAAAATATACAGCTGGTGGCAAGTGTCTTATTCATTCAGCCATTTTCTCAGACA--C | 103283 |
| OY | 6905 | ATTATATTAAGCACTTATATAGTCAGAGCTGAGCTCAAGCTTGATGTCCGAC | 6964 |

[illegible]

| | |
|-----------|--|
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 159492) |
| AUTHORS | Worley, K. C. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| REFERENCE | 3 (bases 1 to 159492) |
| AUTHORS | Worley, K. C. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| COMMENT | On Jul 12, 2002 this sequence version replaced gi:19482237. |
| | ----- Genome Center ----- |
| | Center: Baylor College of Medicine |
| | Center code: BCM |
| | Web site: http://www.hgsc.bcm.tmc.edu/ |
| | Contact: hgsc-help@bcm.tmc.edu |
| | ----- Project Information ----- |
| | Center project name: G0DN |
| | Center clone name: CH230-286017 |
| | ----- Summary Statistics ----- |
| | Sequencing vector: Plasmid |
| | Chemistry: Dye-terminator Big Dye: 100% of reads |
| | Assembly program: Phrap; version 0.990329 |
| | Consensus quality: 124398 bases at least Q40 |
| | Consensus quality: 128059 bases at least Q30 |
| | Consensus quality: 131015 bases at least Q20 |
| | ----- |
| | * NOTE: Estimated insert size may differ from sequence length |
| | * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drfmt_data.html) |
| | * NOTE: This is a 'working draft' sequence. It currently |
| | * consists of 46 contigs. The true order of the pieces |
| | * is not known and their order in this sequence record is |
| | * arbitrary. Gaps between the contigs are represented as |
| | * runs of N, but the exact sizes of the gaps are unknown. |
| | * This record will be updated with the finished sequence |
| | * as soon as it is available and the accession number will |
| | * be preserved. |
| | ----- |
| | 1 1014: contig of 1014 bp in length |
| * | 1015 1114: gap of unknown length |
| * | 1115 2395: contig of 1281 bp in length |
| * | 2396 2495: gap of unknown length |
| * | 2496 3600: contig of 1105 bp in length |
| * | 3601 3700: gap of unknown length |
| * | 3701 5021: contig of 1321 bp in length |
| * | 5022 5121: gap of unknown length |
| * | 5122 6727: contig of 1606 bp in length |
| * | 6728 6827: gap of unknown length |
| * | 6828 8708: contig of 1781 bp in length |
| * | 8609 8709: gap of unknown length |
| * | 8709 10003: contig of 1295 bp in length |
| * | 10004 10103: gap of unknown length |
| * | 10104 11211: contig of 1108 bp in length |
| * | 11212 11311: gap of unknown length |
| * | 11312 12518: contig of 1207 bp in length |
| * | 12519 12618: gap of unknown length |
| * | 12619 13843: contig of 1225 bp in length |
| * | 13844 13943: gap of unknown length |
| * | 13944 15404: contig of 1461 bp in length |
| * | 15405 15505: gap of unknown length |
| * | 15505 17647: contig of 2143 bp in length |
| * | 17648 17747: gap of unknown length |
| * | 17747 19706: contig of 1959 bp in length |
| * | 19707 19806: gap of unknown length |
| * | 19807 21576: contig of 1770 bp in length |
| * | 21577 21676: gap of unknown length |
| * | 21677 22906: contig of 1230 bp in length |
| * | 22907 23006: gap of unknown length |
| * | 23007 25086: contig of 2080 bp in length |
| * | 25087 25186: gap of unknown length |
| * | 25187 26832: contig of 1646 bp in length |

| Query Match | Local Similarity | Score | DB 2; | Length | 159492; |
|---------------|--|--|-----------------|-------------|----------|
| Matches 1950: | Conservative | 0; | Mismatches 664; | Indels 447; | Gaps 25; |
| 168 | ATAGGAGGTTGGCAGTCGTGCACACCTTTAATTCACATATTTCAGAGGACGAGAAACA | 227 | | | |
| db | 48550 | AAAAAGAGGAGGAAAGTGTAGGACGACCTTTAATTCAGGACGACCTTCGAGGACGAGGCA | 48609 | | |
| FEATURES | Source | Location/Qualifiers | | | |
| | 1. 159492 | /organism="Rattus norvegicus" | | | |
| | BASE COUNT | 39083 a 36447 c 36055 g 40464 t 7443 others | | | |
| ORIGIN | | /db_xref="taxon:10116" | | | |
| | | /clone="CH230-286017" | | | |
| 26833 | 26932 | gap of unknown length | | | |
| 28710 | 28710 | contig of 1778 bp in length | | | |
| 28711 | 28810 | gap of unknown length | | | |
| 28811 | 30606 | contig of 1796 bp in length | | | |
| 30607 | 30706 | gap of unknown length | | | |
| 30707 | 33113 | contig of 2407 bp in length | | | |
| 33114 | 33213 | gap of unknown length | | | |
| 33214 | 33661 | contig of 2148 bp in length | | | |
| 35362 | 35461 | gap of unknown length | | | |
| 35462 | 37520 | contig of 2059 bp in length | | | |
| 37521 | 37620 | gap of unknown length | | | |
| 37621 | 40471 | contig of 2851 bp in length | | | |
| 40472 | 40571 | gap of unknown length | | | |
| 40572 | 42584 | contig of 2013 bp in length | | | |
| 42585 | 42684 | gap of unknown length | | | |
| 42685 | 45547 | contig of 2863 bp in length | | | |
| 45548 | 45647 | gap of unknown length | | | |
| 45648 | 47838 | contig of 2191 bp in length | | | |
| 47839 | 47938 | gap of unknown length | | | |
| 47939 | 51229 | contig of 3391 bp in length | | | |
| 51230 | 51329 | gap of unknown length | | | |
| 51330 | 54005 | contig of 2676 bp in length | | | |
| 54006 | 54105 | gap of unknown length | | | |
| 54106 | 56683 | contig of 2578 bp in length | | | |
| 56684 | 56783 | gap of unknown length | | | |
| 56784 | 59361 | contig of 2578 bp in length | | | |
| 59362 | 59461 | gap of unknown length | | | |
| 59462 | 61588 | contig of 2127 bp in length | | | |
| 61589 | 61688 | gap of unknown length | | | |
| 61689 | 65712 | contig of 4024 bp in length | | | |
| 65713 | 65812 | gap of unknown length | | | |
| 65813 | 69339 | contig of 3527 bp in length | | | |
| 69340 | 69439 | gap of unknown length | | | |
| 69440 | 72570 | contig of 3131 bp in length | | | |
| 72571 | 72670 | gap of unknown length | | | |
| 72671 | 76898 | contig of 4228 bp in length | | | |
| 76899 | 76998 | gap of unknown length | | | |
| 76999 | 80760 | contig of 3762 bp in length | | | |
| 80761 | 80860 | gap of unknown length | | | |
| 80861 | 86313 | contig of 5453 bp in length | | | |
| 86314 | 86413 | gap of unknown length | | | |
| 86414 | 92231 | contig of 5818 bp in length | | | |
| 92232 | 92331 | gap of unknown length | | | |
| 92332 | 96992 | contig of 7361 bp in length | | | |
| 96993 | 99792 | gap of unknown length | | | |
| 99793 | 106178 | contig of 6586 bp in length | | | |
| 106179 | 106278 | gap of unknown length | | | |
| 106279 | 114304 | contig of 8026 bp in length | | | |
| 114305 | 114405 | gap of unknown length | | | |
| 114405 | 122112 | contig of 7708 bp in length | | | |
| 122113 | 122212 | gap of unknown length | | | |
| 122213 | 129642 | contig of 7430 bp in length | | | |
| 129643 | 129742 | gap of unknown length | | | |
| 129743 | 138684 | contig of 8942 bp in length | | | |
| 138685 | 138784 | gap of unknown length | | | |
| 138785 | 149662 | contig of 10878 bp in length | | | |
| 149663 | 149762 | gap of unknown length | | | |
| 149763 | 159492 | contig of 9730 bp in length. | | | |

[illegible][illegible]


```

Db 50618 ----- 50617
Oy 2287 GAAGACAGCTGCAGAGCTACAGTGTACTAGTATATCTATATATTAATCTTTTAA 2346
Db 50618 -----NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Oy 2347 AAAATGAGAGGATCTGAGACCTCAAAAGATTTATGACAGTACTACGGGTGAT 2406
Db 50636 NCAGTGGAGAGAGAGCTCAGACACTCAAAATAGATTATGGCAGCTACTCATGGGTGAT 50715
Oy 2407 TATCTATCCTGGAGTCTTTTCCCTTTCCGCTTGGCTTGCACACTGGTGGAGAGAGCCCTT 2466
Db 50716 TATCTATCCTGGAGTCTTTTCCGCTTGGCTTGTCTGTACTGGGTGGAGACAGCCTCTT 50775
Oy 2467 TTCTATTCACAGAGACGGGTCTACTATTATTTTGAACAACAGCAGCTGCAATGTTT 2526
Db 50776 TCCATTTCACAAAGATAGAGGCTACATTATTTCTAAACAAACAGCAGCTGCAATGTTT 50835
Oy 2527 ACTGTCCTTGTGATGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2586
Db 50836 ACTGTCCTCAGGAGCTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 50868
Oy 2587 ACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGAGAGTCAAGAGA 2646
Db 50869 -----CACACACACATTCAGTCTCCAGAGCTTTGGAGAGTCAAGAGA 50915
Oy 2647 GGCTGCCCTCAACACAGATCTTCTATCTTTCCCTCCTTAAGAGACCGATTCAGAGTG 2706
Db 50916 GGCTGCCCTCAACACAGATCTTCTATCTTTCCCTCCTTAAGAGACCGATTCAGAGTG 50973
Oy 2707 GCAGAGATCTACAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2766
Db 50974 GCAGAGATCTACAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51033
Oy 2767 CTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2826
Db 51034 CTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51093
Oy 2827 TTCCGAGC-----AGAGGCGAGCGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
Db 51094 GCTCCTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51153
Oy 2881 GCTCTAAAGAGTGAAGGAGTCTGCTGCTCAGTCAAGAGAGGAGGAGGAGGAGGAGGAG 2940
Db 51154 ACTCTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51209
Oy 2941 GCGCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000
Db 51210 GCGCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 51269
Oy 3001 CTGCTGTGCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3060
Db 51270 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 51329
Oy 3061 C 3061
Db 51330 C 51330

```

```

RESULT 5
ACLI15173/ 159492 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-286017, *** SEQUENCING IN PROGRESS
DEFINITION *** 46 unordered pieces.
ACCESSION ACI15173
VERSION ACI15173.2 GI:21738030
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 159492)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,V., Benton,J., Blincoe,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.J., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frenzt,P.,
Gabisi,A., Gao,Y., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Huijck,S., Hune,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtharge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogum,B., Okunolu,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Plimus,E., Pu,L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rojle,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tamey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Welstock,G. and Gibbs,R.
Direct Submission
2 (bases 1 to 159492)
Worley,K.C.
Direct Submission
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 159492)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:19482237.
-----
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GCDN
Center clone name: CH230-286017
-----
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 124598 bases at least Q40
Consensus quality: 128059 bases at least Q30
Consensus quality: 131015 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length

```


Db 28031 TGGCTTCGTTCTCCGGGTGACAAAGTGTCTCCAGGG-----ATGGGGACC 27984
Qy 5150 AGGGGTGAGCAATGAGTTTGAAGTGGAGACATCCAGCCAGGGTCTTGGCGGAA 5209
Db 27983 AGGGGTGAGTCAAGTGTGAAGTGAAGCAATATCTACCCAGGGTCTTGGCGAG 27924
Qy 5210 GCTAAAGAAAGAGACCCCTTACCCCTCTGAAGTTTGGGAGACAGAGAGCTGAGG 5269
Db 27923 ACTGAAGATAGAG -CCTCCAGCCCTCTTGAAGTTTGGGAGACCTGGGAGCTGAGG 27865
Qy 5270 AGATCTTCTAAGGTGAGAGAGAGTATCT -GCTCTGACCAACATGCTAGAGAGAG 5328
Db 27864 ATGTCCTCTGAGAGAGAGAGTGTCTCTCTGACCAAGATGAGGTGCAAGAG 27805
Qy 5329 CA-----GTTGACCAGTTTACCCCTCAAGACAGCATCCCTCTTG 5370
Db 27804 CACTTGAAGATACCCAGAGGGTGTGGCCGGCTGCTCAAGAGCCAGCATCCACCTTG 27745
Qy 5371 GCTCAAGAGAGCTGGGCGCCCTTCTGTTTGAAGTCTTACTTTTCTCAGAGAGAGCA 5430
Db 27744 GCCCTAAGAGAGCTGTGCCCCCTTTCGGTTTAAAGATCTCAC -TTTCTTCAAGAGAGACA 27686
Qy 5431 GCAAGCCCTTGTCCCTCCCTGTTGTCAATTAACACCCCTGTGTACATTAATTACTTAT 5490
Db 27685 GCAAGCCCTTGT -CCCCGCCAAGACAGTCAATTAACATCCCTGTGTCAACATTAATTAT 27627
Qy 5491 TTTACTGTCATTTGCTTCAAGAGAGTCCATCTGTGTAGACCTGTGCTCTTAACACCA 5550
Db 27626 TTTTCTGTCATTTGCTTCAAGAGAGTCCATTTGGTATACCTTCCCTTGTCCACCA 27567
Qy 5551 GGTATGGCCCCACATTTCTCACCAGAGAGTGCAGAGAGAGCCTTTACAGAAAGGGTAC 5610
Db 27566 GCTATGGCCCCACCTTCTCACCAGAG -AGTGTGCAAAAGAGAGCCTTTACAGAAAGGTAC 27508
Qy 5611 AGTAAACAAGATGGCCAGAAATTAACAAACAACTACTATCTTTTACCCAAATTTGGTTT 5670
Db 27507 AGTAAACAAGAGCGCCAGAAATTAACAAATACTGTCTTTTGTACTAG -TT 27454
Qy 5671 GCTAACCAGAGAGGGGTGTGTGAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5730
Db 27453 GCTAACCAGAG----- 27442
Qy 5731 GT 5790
Db 27441 -----GTGAGT 27332
Qy 5791 TGATATTGGCCCATGCGCAACAGGGGTATTGGGAGAGTC-----AGGCTTCTGC 5841
Db 27391 TGATATTGGGCCCATGCGCAATAGGGGTATCAGGTAGAGTCAGGCGCTGTAAGGCTTCTGC 27332
Qy 5842 AAACAGATAGCTTCCCAAGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5901
Db 27331 AAACAGACAGAGTGGCCCAAGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 27272
Qy 5902 AGAGTGTGACAAATCATCAAGATTAAGCACCCTGTGGGCTCAGAAAGAGGAGTTTACA 5961
Db 27271 AGAGTGTGACAA-----CAACAATTAAGCACCCTGTGAGGCTCAG-----GAGTTTACA 27220
Qy 5962 AGAGTAAAGGCCAAGCCATTATTATTCAAAGACATGACTCAAAATCAAGTGAAGAG 6021
Db 27219 AGAGTAAAGGCCAAGCCATTATTATTCAAAGACATGACTCAAAATCAAGTGAAGAG 27160
Qy 6022 AGATTAGCTGAGAGATGGGGCTGTGATGTGGGACACCTGACCTTGACCTTATTATAGTCA 6081
Db 27159 AGATTAGCTGAGAGATGGGGCTGTGATGTGGGATTAACCTTGTGCCCTTGTAGTCA 27100
Qy 6082 CTAGGCCAAGAGAGATGACAGAGGGTACTGCTACTACGCTTGGGAGGAGCAGCT 6141
Db 27099 CTAGGCCAAGAGAGATGACAGAGGGTACTGCTACTACGCTTGGGAGGAGCAGAT 27040
Qy 6142 GGAGAAATGGTACCTTCATCTGATGAGAGGGCTGAGACCAACAGCAGTACAACTGTTC 6201
Db 27039 AGAGAAATGG-----TGGTGAGAGGGTTGAGAAATGACAGCTGCAACTGTTC 26992

Qy 6202 C---CTGTGTCTCATGCCAGATTCCTGGCCAGTTTCAAGAGACTAAGACTATCT 6258
Db 26991 CTCATGTGCTCATGCTGGGATTTCCAGCAGTTTCAAGAGACCAGACTATCTCN 26932
Qy 6259 GGTGAAACAAGTATCCAGCCCTTAAGCCCATTTTGGTCTAATTAATCAAGAACCCCT 6318
Db 26931 NNN 26872
Qy 6319 GGGATGCAAGCTCTGACACAGCAGAGCTTTTAAAAAGCT 6359
Db 26871 NNTT 26831

RESULT 6
AC119514/c
LOCUS
DEFINITION
AC119514 Rattus norvegicus clone CH230-407A8, *** SEQUENCING IN PROGRESS
VERSION
AC119514.3 GI:21746577
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS
1 (bases 1 to 80578)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbacia,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnah,D.,
Bouck,C., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,D., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinb,H.,
Douthwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H.,Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlek,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Holjk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegod,H.,
Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maneshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Medford,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkenko,S., Ogun,M., Okunodu,G.,
Oregunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L., Ruiz,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Rulles,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostlati,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,N., Stone,H.,
Sutton,A., Syarik,A., Tabor,P., Tamerlis,A., Tamerlis,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Woodem,C., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 80578)
REFERENCE
AUTHORS
Worley,K.C.
TITLE
Direct Submission

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Db | 59121 | TACATCGCAGGTCGAACACCTCAGCACAATTAATAATTAATAATTAATAATTAATTC | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

| | |
|-----------|---|
| REFERENCE | REUTER M., RICHARDS S., BIGGS F., RIVES C., RODKEY T., ROJAS A., ROSE M., ROSE R., RUIZ S.J., SANDERS M., SAVERY G., SCHERER S., SCOTT G., SHATSMAN S., SHEN H., SHELLEY J., SHVARTSBERG A., SLOSSON I., SITTER C.D., SMAJS D., SODERGREN E., SONG X.-Z., SOTELLE R., SOSA J., STEINLE M., STRONG R., SUTTON A., SVATEK A., TABOR P., TAYLOR C., TAYLOR T., THOMAS N., THOMAS S., TINGEY A., TREJOS Z., USMANT K., VALAS R., VERA V., VILLASANA D., WALDON L., WALKER B., WANG J., WANG Q., WANG S., WARREN J., WARTEN R., WEI X., WHITE F., WILLIAMS G., WILLSON R., WLECZYK R., WOODEN H., WORLEY K., WRIGHT R., WRIGHT R., WU J., YAKUB S., YEN J., YOON L., YOON V., YU F., ZHANG J., ZHOU J., ZHOU X., ZHAO S., DUNN D., VON NIEDERHAUSEN A., WEISS R., SMITH D.R., HOLT R.A., SMITH H.O., WEINSTOCK G. and GIBBS R.A. |
| TITLE | Direct Submission |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 181772) |
| AUTHORS | Morley,K.C. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| REFERENCE | 3 (bases 1 to 181772) |
| AUTHORS | Rat Genome Sequencing Consortium. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA |
| COMMENT | On Aug 3, 2002 this sequence version replaced gi:21700341. |
| | ----- Genome Center ----- |
| | Center: Baylor College of Medicine |
| | Center code: BCM |
| | Web site: http://www.hgsc.bcm.tmc.edu/ |
| | Contact: hgsc-help@bcm.tmc.edu |
| | Project Information |
| | ----- |
| | Center project name: CH230-2B3 |
| | Center clone name: CH230-2B3 |
| | ----- |
| | Summary Statistics |
| | Sequencing vector: Plasmid. |
| | Chemistry: Dye-terminator Big Dye: 100% of reads |
| | Assembly program: Phrap: version 0.990329 |
| | Consensus quality: 135179 bases at least Q40 |
| | Consensus quality: 140704 bases at least Q30 |
| | Consensus quality: 145423 bases at least Q20 |
| | ----- |
| | * NOTE: Estimated insert size may differ from sequence length |
| | * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html) |
| | * NOTE: This is a 'working draft' sequence. It currently |
| | * consists of 55 contigs. The true order of the pieces |
| | * is not known and their order in this sequence record is |
| | * arbitrary. Gaps between the contigs are represented as |
| | * runs of N, but the exact sizes of the gaps are unknown. |
| | * This record will be updated with the finished sequence |
| | * as soon as it is available and the accession number will |
| | * be preserved. |
| | 1 |
| | 1211: contig of 1210 bp in length |
| * | 1211: gap of unknown length |
| * | 1311: contig of 1598 bp in length |
| * | 2908: gap of unknown length |
| * | 3009: contig of 1597 bp in length |
| * | 4605: gap of unknown length |
| * | 4705: contig of 1702 bp in length |
| * | 6407: gap of unknown length |
| * | 6508: contig of 1253 bp in length |
| * | 7760: gap of unknown length |
| * | 7761: contig of 1428 bp in length |
| * | 9288: gap of unknown length |
| * | 9289: gap of unknown length |
| * | 9389: contig of 1027 bp in length |
| * | 10416: gap of unknown length |
| * | 10516: contig of 1608 bp in length |
| * | 12123: gap of unknown length |
| * | 12124: contig of 1129 bp in length |
| * | 12352: gap of unknown length |
| * | 13353: contig of 145423 bases at least Q20 |
| * | 13452: gap of unknown length |
| * | 13453: contig of 1703 bp in length |

| | | | |
|---|--------|---------|-----------------------------|
| * | 15156 | 15255: | gap of unknown length |
| * | 15256 | 16618: | contig of 1363 bp in length |
| * | 16619 | 16718: | gap of unknown length |
| * | 16719 | 18622: | contig of 1904 bp in length |
| * | 18623 | 18722: | gap of unknown length |
| * | 18723 | 19829: | contig of 1107 bp in length |
| * | 19830 | 19929: | gap of unknown length |
| * | 19930 | 21969: | contig of 2040 bp in length |
| * | 21970 | 22069: | gap of unknown length |
| * | 22070 | 22915: | contig of 1846 bp in length |
| * | 22916 | 24015: | gap of unknown length |
| * | 24016 | 26059: | contig of 2004 bp in length |
| * | 26060 | 26159: | gap of unknown length |
| * | 26160 | 28222: | contig of 2063 bp in length |
| * | 28223 | 28322: | gap of unknown length |
| * | 28323 | 30399: | contig of 2077 bp in length |
| * | 30400 | 30499: | gap of unknown length |
| * | 30500 | 32655: | contig of 2156 bp in length |
| * | 32656 | 32755: | gap of unknown length |
| * | 32756 | 33854: | contig of 1099 bp in length |
| * | 33855 | 33954: | gap of unknown length |
| * | 33955 | 35237: | contig of 1283 bp in length |
| * | 35238 | 35337: | gap of unknown length |
| * | 35338 | 36771: | contig of 1434 bp in length |
| * | 36772 | 36871: | gap of unknown length |
| * | 36872 | 38777: | contig of 1906 bp in length |
| * | 38778 | 38877: | gap of unknown length |
| * | 38878 | 40669: | contig of 1792 bp in length |
| * | 40670 | 40769: | gap of unknown length |
| * | 40770 | 43169: | contig of 2400 bp in length |
| * | 43170 | 43269: | gap of unknown length |
| * | 43370 | 45568: | contig of 2299 bp in length |
| * | 45569 | 45668: | gap of unknown length |
| * | 45669 | 47858: | contig of 2190 bp in length |
| * | 47859 | 47958: | gap of unknown length |
| * | 47959 | 49767: | contig of 1809 bp in length |
| * | 49768 | 49867: | gap of unknown length |
| * | 49868 | 52667: | contig of 2700 bp in length |
| * | 52668 | 52667: | gap of unknown length |
| * | 52669 | 55159: | contig of 2492 bp in length |
| * | 55160 | 55259: | gap of unknown length |
| * | 55260 | 57887: | contig of 2628 bp in length |
| * | 57888 | 57987: | gap of unknown length |
| * | 57988 | 60041: | contig of 2054 bp in length |
| * | 60042 | 60141: | gap of unknown length |
| * | 60142 | 62866: | contig of 2725 bp in length |
| * | 62867 | 62966: | gap of unknown length |
| * | 62967 | 66014: | contig of 3048 bp in length |
| * | 66015 | 66114: | gap of unknown length |
| * | 66115 | 68172: | contig of 2058 bp in length |
| * | 68173 | 68272: | gap of unknown length |
| * | 68273 | 72352: | contig of 4080 bp in length |
| * | 72353 | 72452: | gap of unknown length |
| * | 72453 | 76168: | contig of 3716 bp in length |
| * | 76169 | 76268: | gap of unknown length |
| * | 76269 | 79795: | contig of 3521 bp in length |
| * | 79796 | 79895: | gap of unknown length |
| * | 79896 | 84438: | contig of 4543 bp in length |
| * | 84439 | 84538: | gap of unknown length |
| * | 84539 | 88636: | contig of 4118 bp in length |
| * | 88637 | 88756: | gap of unknown length |
| * | 88757 | 93275: | contig of 4519 bp in length |
| * | 93276 | 93375: | gap of unknown length |
| * | 93376 | 98043: | contig of 4668 bp in length |
| * | 98044 | 98143: | gap of unknown length |
| * | 98144 | 103024: | contig of 4881 bp in length |
| * | 103025 | 103124: | gap of unknown length |
| * | 103125 | 107830: | contig of 4706 bp in length |
| * | 107831 | 107930: | gap of unknown length |
| * | 107931 | 114501: | contig of 6571 bp in length |
| * | 114502 | 114601: | gap of unknown length |
| * | 114602 | 119361: | contig of 4760 bp in length |
| * | 119362 | 119361: | gap of unknown length |

| Query Match | 2.1%: Score 188.6; DB 2; Length 181772; |
|--|--|
| Best Local Similarity | 62.4%: Pred. No. 2.4e-43; |
| Matches 384; Conservative | 0; Mismatches 214; Indels 17; Gaps 5; |
| QY 1799 ATATATGTTACATATGATTAATTAATTAATTAATTCCTTTAAAGATTAATGTTGGAGG 1858 | 119462 125323: contig of 5862 bp in length |
| Db 78941 ATTTTATGTAAGCCACAAAGAAATGATATCATTTTGCCACAAATATGTTGGAGACTG 79000 | * 125324 125423: gap of unknown length |
| QY 1859 GATAGATGGCTCAGCTTCCAGAGACACTGCGTGGCTTGCAGAGAGACTACATTCAGTTC 1918 | * 125424 134075: contig of 8652 bp in length |
| Db 79001 GATAGAGAGCTAAGAGGCTTAAGAGCAGTGGCGCTTTGCAGAGAGACCTGCTTCAGATTC 79060 | * 134076 134175: gap of unknown length |
| QY 1919 CCAAGACTATATGAGGCTGCACAGCACTCTTAATCCAGTTCAGAGAGGTTCCACAC 1978 | * 134176 138728: contig of 4553 bp in length |
| Db 79061 CCAAGACCCACAGCGTACTCACACACATTTTAATCTAGTCCAGAGGGCCCCAAGACC 79120 | * 138729 138828: gap of unknown length |
| QY 1979 CTCTCTGGCCCTCCACAGGCCACCATATCATAGTACAGACATATCATGACGGCAAAACA 2038 | * 138829 142482: contig of 3654 bp in length |
| Db 79121 CTCTTCTGGCATCTGTGTGCACACATGCACACGGTATACAGGCATTCAGG-AAAACA 79179 | |
| QY 2039 CCATATACACATTAATTAATTAAGCAACTTAAAGGTGCATGTGTGTAACATTTGTG 2098 | |
| Db 79180 CTCACACATACGAACTGTGATATCATTTTTTTTAACTTATGTGTGTTGCTACAGATATGGA 79239 | |
| QY 2099 CTATACACTGTGATGTGAAGACA---TGTACACGCCACACACT---GAAAGGGATCTGG 2152 | |
| Db 79240 TGCAGCCACAGACCATGATGATTAAGAGCACTGACTCTCTTCCGAAAGGAGTCTGA 79299 | |
| QY 2153 GCGTGGAGAGATGGCTCAGCGGTTAAGAGCACTGACTCTCTTCCGAAAGGAGTCTGA 2212 | |
| Db 79300 GCGTGGAGAAATGGCTCAGTATTAATAGCACTGAGTGCCTGCC---AGAGTCTCGA 79355 | |
| QY 2213 GTTCAATCTCTAGCAACCATGTTGGCTCACACACCATTCATTAATGATCTGCACCCCT 2272 | |
| Db 79356 GTTCAATCTCCACAGCAACCATGTTGGCTCACACACCATTCGTAATGAGATCTGTAGCCCT 79415 | |
| QY 2273 CTCTGGTGCATCTCTGAAGACAGCTGAGAGGTACAGTGTACTATTAATTAATTAATA 2332 | |
| Db 79416 CTCTGGTGTCTCTGAAGACAGCTACAG-----TGTACTCATTAATTAATTAATAA 79469 | |
| QY 2333 AATCTTTTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTATGAGCAGT 2392 | |
| Db 79470 ATCTTTTAAAAAAGAGAAATATTATATCTTCTCTGCATGTGACCTTAGGTTT 79529 | |
| QY 2393 GACTCAGCGGTGATT 2407 | |
| Db 79530 TTTTAAAGATTATT 79544 | |
| RESULT 8 | |
| AC128412/c | 194630 bp DNA linear HTG 19-JUL-2002 |
| LOCUS | Rattus norvegicus clone CH230-152K9, *** SEQUENCING IN PROGRESS |
| DEFINITION | ***, 59 unordered pieces. |
| ACCESSION | AC128412 |
| VERSION | AC128412.1 GI:21909085 |
| KEYWORDS | HTG; HTGS_PHASE1. |
| SOURCE | Rattus norvegicus. |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; |
| REFERENCE | Rattus. |
| AUTHORS | 1 (bases 1 to 194630) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alisbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., |

Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonaldi, D.,
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eathat, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Krstovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lousaged, H.,
 Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Monabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunolu, G.,
 Orangun, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Sytek, A., Tabor, P., Tameris, A., Tameris, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,
 Wang, Q., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Woodem, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 194630)
 Morley, K.C.

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GYO
 Center clone name: CH230-152K9

----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 135641 bases at least Q40
 Consensus quality: 141731 bases at least Q30
 Consensus quality: 146942 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_difft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 59 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1793: contig of 1793 bp in length

1794 1893: gap of unknown length
 1894 3132: contig of 1239 bp in length
 3133 3232: gap of unknown length
 3233 4328: contig of 1096 bp in length
 4329 4428: gap of unknown length
 4429 5580: contig of 1152 bp in length
 5581 5681: gap of unknown length
 5681 7470: contig of 1790 bp in length
 7471 7570: gap of unknown length
 7571 9126: contig of 1556 bp in length
 9127 9226: gap of unknown length
 9227 10595: contig of 1369 bp in length
 10596 10695: gap of unknown length
 10696 11883: contig of 1188 bp in length
 11884 11983: gap of unknown length
 11984 13414: contig of 1431 bp in length
 13415 14925: contig of 1411 bp in length
 14926 15025: gap of unknown length
 15026 16379: contig of 1354 bp in length
 16380 16479: gap of unknown length
 16480 17657: contig of 1178 bp in length
 17658 17757: gap of unknown length
 17758 18831: contig of 1074 bp in length
 18832 18931: gap of unknown length
 18932 20123: contig of 1192 bp in length
 20124 20223: gap of unknown length
 20224 21898: contig of 1675 bp in length
 21899 23422: gap of unknown length
 23423 23523: contig of 1424 bp in length
 23523 25401: gap of unknown length
 25402 25501: gap of unknown length
 25502 27167: contig of 1666 bp in length
 27168 27267: gap of unknown length
 27268 28730: contig of 1463 bp in length
 28731 30823: gap of unknown length
 30824 30923: contig of 1993 bp in length
 30924 32988: gap of unknown length
 32989 33088: gap of unknown length
 33089 35579: contig of 2491 bp in length
 35580 35679: gap of unknown length
 35680 37193: contig of 1514 bp in length
 37194 37293: gap of unknown length
 37294 38574: contig of 1281 bp in length
 38575 38674: gap of unknown length
 38675 40473: contig of 1799 bp in length
 40474 40573: gap of unknown length
 40574 42832: contig of 2259 bp in length
 42833 42932: gap of unknown length
 42933 44836: contig of 1904 bp in length
 44837 44936: gap of unknown length
 44937 46922: contig of 1866 bp in length
 46923 47022: gap of unknown length
 47023 48984: contig of 1662 bp in length
 48985 49084: gap of unknown length
 49085 52798: contig of 3714 bp in length
 52799 52898: gap of unknown length
 52899 55863: contig of 2965 bp in length
 55864 55963: gap of unknown length
 55964 57728: contig of 1765 bp in length
 57729 57828: gap of unknown length
 57829 59470: contig of 1642 bp in length
 59471 62556: gap of unknown length
 62557 62657: gap of unknown length
 62657 65039: contig of 2383 bp in length
 65040 65139: gap of unknown length
 65140 67858: contig of 2719 bp in length
 67859 70147: gap of unknown length
 70148 70247: contig of 2189 bp in length

| | | | | | | | | |
|---|--------|---------|--------|----|---------|--------|----|--------|
| * | 70248 | 72929: | contlg | of | 2682 | bp | in | length |
| * | 72930 | 73029: | gap | of | unknown | length | | |
| * | 73030 | 76076: | contlg | of | 3047 | bp | in | length |
| * | 76077 | 76176: | gap | of | unknown | length | | |
| * | 76177 | 79311: | contlg | of | 3135 | bp | in | length |
| * | 79312 | 79411: | gap | of | unknown | length | | |
| * | 82901 | 82900: | contlg | of | 3489 | bp | in | length |
| * | 83001 | 83000: | gap | of | unknown | length | | |
| * | 87435 | 87434: | contlg | of | 4434 | bp | in | length |
| * | 87553 | 87534: | gap | of | unknown | length | | |
| * | 92931 | 92930: | contlg | of | 5396 | bp | in | length |
| * | 93031 | 93030: | gap | of | unknown | length | | |
| * | 96310 | 96209: | contlg | of | 3179 | bp | in | length |
| * | 96320 | 96309: | gap | of | unknown | length | | |
| * | 99070: | 99070: | contlg | of | 2761 | bp | in | length |
| * | 99071: | 99170: | gap | of | unknown | length | | |
| * | 99171 | 102188: | contlg | of | 3018 | bp | in | length |
| * | 102189 | 102288: | gap | of | unknown | length | | |
| * | 102289 | 105629: | contlg | of | 3341 | bp | in | length |
| * | 105630 | 105729: | gap | of | unknown | length | | |
| * | 105730 | 111114: | contlg | of | 5385 | bp | in | length |
| * | 111115 | 111214: | gap | of | unknown | length | | |
| * | 111215 | 114351: | contlg | of | 3137 | bp | in | length |
| * | 114352 | 114451: | gap | of | unknown | length | | |
| * | 114452 | 119170: | contlg | of | 4719 | bp | in | length |
| * | 119171 | 119270: | gap | of | unknown | length | | |
| * | 119271 | 123696: | contlg | of | 4426 | bp | in | length |
| * | 123697 | 123796: | gap | of | unknown | length | | |
| * | 123797 | 133294: | contlg | of | 9498 | bp | in | length |
| * | 133295 | 133294: | gap | of | unknown | length | | |
| * | 133395 | 141172: | contlg | of | 7778 | bp | in | length |
| * | 141173 | 141172: | gap | of | unknown | length | | |
| * | 141273 | 148661: | contlg | of | 7389 | bp | in | length |
| * | 148662 | 148761: | gap | of | unknown | length | | |
| * | 148762 | 157010: | contlg | of | 8249 | bp | in | length |
| * | 157011 | 157110: | gap | of | unknown | length | | |
| * | 157111 | 166499: | contlg | of | 9359 | bp | in | length |
| * | 166470 | 166569: | gap | of | unknown | length | | |
| * | 166570 | 175385: | contlg | of | 8816 | bp | in | length |
| * | 175386 | 175485: | gap | of | unknown | length | | |

| Query Match | Similarity | 2.1% | Score | 188.6 | DB.2 | Length | 194630 |
|-------------|---|--------------|-------|------------|--------|--------|--------|
| Best Local | Similarity | 67.4% | Pred. | No. | 2.4e+3 | | |
| Matches | 359 | Conservative | 0 | Mismatches | 134 | Indels | 40 |
| | | | | | | | Gaps |
| | | | | | | | 5 |
| QY 1846 | ATATGTTGGGAGGAGAGATGGCTCAGCTTCCAGAGACCTTGCTGCTTGCAGAGAGAC | 1905 | | | | | |
| Db 74564 | ATTTTGGGGGGCTGGAGAGATGGTTCAGTGGTTAGGATCGACTGGTGGCTTCCAGAGGAT | 74505 | | | | | |
| QY 1906 | CTAGATTGAGTTCCAGGACCTATATGTGGGTCCACAGGACCTCTGTAATTCAGTTCCAG | 1965 | | | | | |
| Db 74504 | GTGGGCTACATTTTAGTACCCATATATGAGGCTCACAAACCTTTGTAACTCCAGTTCCAT | 74445 | | | | | |
| QY 1966 | AGGTTTCCACACACCTCTTGTGGGCTCCACAGGACACACACACACACACACACACAC | 2021 | | | | | |
| Db 74444 | GGGATCCGATACCATCTCTGTGACTTCCCTGGGGCCCGCAGGACATGTCACACAGACA | 74385 | | | | | |
| QY 2022 | TACATGCAGGCAAAACCCATACACACATAA-ATTAATTAAGAAACCTTAAAGGTGCAT | 2080 | | | | | |
| Db 74384 | TACATGCAGGCTCAAAACATCAGCACACATAAATAAATAATCAATAATATATGT | 74325 | | | | | |
| QY 2081 | GGTGTGGTAAACATTTGTGTTACACATGCTGATTTGAAGCATGTACAACGCACACACCTA | 2140 | | | | | |
| Db 74324 | GTGTATGTGTGTGTATTTACACACACACCTATATATTTATTTTTAGTCACAGCTTTTCTGT | 74265 | | | | | |
| QY 2141 | AGAG-----GGATTTGGGGCTGGAGAGATGGCTCAGCGGTTA | 2177 | | | | | |
| Db 74264 | GTAGCCCTGTGTTTCTTAAACTCACTCTGTAGGGGGCTGGAGAGGTGGCTCAGCGGTTA | 74205 | | | | | |
| QY 2178 | AGAGCAGTACTGCTCTTCCAGAGAGGCTCTGAGTTCAATTCCTAGCAACCATGTGT | 2237 | | | | | |
| Db 74204 | AGAGCAGTACTGCTCTTCC-----AGAGTCTCTGAGTTCAATTCCTCAGCAACCATGTGT | 74149 | | | | | |

| | | | |
|------------|--|---|-------|
| Oy | 2238 | GGCCTACACAACCATCCATATAATGAGATCTGACACCCTTCTGTGGATCTGAAGACAGCTG | 2297 |
| | | | |
| Db | 74148 | GGCTTCACAACCATCTGTATATGGAGTCATGACCCCTTTCTGTGTGAAGAC---- | 74094 |
| Oy | 2298 | CAGACTCTACAGTGATCTAGATATATCTCAATAAATAAACTTTTTTAATAAAAAA | 2350 |
| | | | |
| Db | 74093 | ---GGCTACAGCTCCTACTTATATATATATAATCAATAAATCTTTTTTAAAAAAA | 74044 |
| RESULT 9 | | | |
| AC013320 | | | |
| LOCUS | | | |
| DEFINITION | AC013320 | 149425 bp DNA linear HTG 18-NOV-2000 | |
| | Homo sapiens chromosome 2 clone RP11-36003 map 2, WORKING DRAFT | | |
| SEQUENCE | | 19 unordered pieces. | |
| ACCESSION | AC013320 | | |
| VERSION | AC013320.7 | GI:11225433 | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. | | |
| SOURCE | Homo sapiens. | | |
| ORGANISM | Homo sapiens | | |
| | Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | |
| REFERENCE | 1 (bases 1 to 149425) | | |
| AUTHORS | Birtren,B., Linton,L., Nusbaum,C. and Lander,E. | | |
| TITLE | Homo sapiens chromosome 2, clone Rpl1-36003 | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 149425) | | |
| AUTHORS | Birtren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., | | |

TITLE Direct Submission
JOURNAL Submitted (36-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 330 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 18, 2000 this sequence version replaced gi:10864221.

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3776
Center clone name: 360_O_3

----- Summary Statistics
Sequencing vector: M13: M77815; 12% of reads
Sequencing vector: Plasmid: n/a; 88% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 140867 bases at least Q40
Consensus quality: 144879 bases at least Q30
Consensus quality: 146417 bases at least Q20
Insert size: 228000; agarose-fp
Insert size: 147625; sum-of-ctrls
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 5.5 in Q20 ba.

NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```



```

* as soon as it is available and the accession number will
* be preserved.
1
8145 8244: contig of 8144 bp in length
8245 9350: contig of 1106 bp in length
9351 9450: gap of 100 bp
9451 11335: contig of 1885 bp in length
11336 11435: gap of 100 bp
11436 14325: contig of 2890 bp in length
14326 14425: gap of 100 bp
14426 17305: contig of 2880 bp in length
17306 17405: gap of 100 bp
17406 20165: contig of 2761 bp in length
20167 20266: gap of 100 bp
20267 23474: contig of 3208 bp in length
23475 23574: gap of 100 bp
23575 26403: contig of 2829 bp in length
26404 26503: gap of 100 bp
26504 29899: contig of 3396 bp in length
29900 29999: gap of 100 bp
30000 34929: contig of 4930 bp in length
34930 35029: gap of 100 bp
35030 41463: contig of 6434 bp in length
41464 41563: gap of 100 bp
41564 47890: contig of 6327 bp in length
47891 47990: gap of 100 bp
47991 56092: contig of 8102 bp in length
56093 56192: gap of 100 bp
56193 63063: contig of 6871 bp in length
63064 63163: gap of 100 bp
63164 70819: contig of 7656 bp in length
70820 70919: gap of 100 bp
70920 82482: contig of 11563 bp in length
82483 82582: gap of 100 bp
82583 98645: contig of 16063 bp in length
98646 98745: gap of 100 bp
98746 117782: contig of 19037 bp in length
117783 117883 149425: gap of 100 bp
117883 149425: contig of 31543 bp in length.
Location/Qualifiers
1. 149425
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-36003"
/clone_lib="RPCT-11 Human Male BAC"
1. 8144
/note="assembly-fragment
clone_end:SP6
vector_side:left"
misc_feature
8245..9350
/note="assembly-fragment"
9451..11335
/note="assembly-fragment"
11436..14325
/note="assembly-fragment"
14426..17305
/note="assembly-fragment"
17406..20166
/note="assembly-fragment"
20267..23474
/note="assembly-fragment"
23575..26403
/note="assembly-fragment"
26504..29899
/note="assembly-fragment"
30000..34929
/note="assembly-fragment"
35030..41463
/note="assembly-fragment"
41564..47890
/note="assembly-fragment"

```

| | | |
|---------------------------|--|---------------------------|
| misc_feature | 47991..56092 | /note="assembly_fragment" |
| misc_feature | 56193..63063 | /note="assembly_fragment" |
| misc_feature | 63164..70819 | /note="assembly_fragment" |
| misc_feature | 70920..82482 | /note="assembly_fragment" |
| misc_feature | 82583..98645 | /note="assembly_fragment" |
| misc_feature | 98746..117782 | /note="assembly_fragment" |
| misc_feature | 117883..149425 | /note="assembly_fragment" |
| misc_feature | clone_end:77 | vector_side:right" |
| BASE COUNT | 36100 a 37729 c 37628 g 36165 t 1803 others | |
| ORIGIN | | |
| Query Match | 2.1%; Score 188; DB 2; Length 149425; | |
| Best Local Similarity | 67.4%; Pred. No. 3.4e-43; | |
| Matches 507; Conservative | 0; Mismatches 190; Indels 55; Gaps 15; | |
| 3238 | AAAGCCAAAGGCGGGGACACGGCAGCTCAGGCTGTGCGCTTGTGGCTTACGTT | 3237 |
| DB | 44776 AAAGCCAAAGGCGGGGACACA--TCAGCTCTCATTTGTACAGCTG-TGAGAGGCTTGGGCT | 44832 |
| QY | 3298 CCCAGGAGCCTGG--GCACCTACTTCCCAACCCGCCATCCATTCCTCTGGGCCCTAT | 3355 |
| DB | 44833 CCCAGGAGCCTGGCAGCAGCTCTCCGCCACCCCTCAC----CTCTTGGGCCCTAT | 44887 |
| QY | 3356 CTTCCTTATATAGTGAAGAAGATTCTTGGGGGGGGGTGGTGAAGACAAGATC | 3415 |
| DB | 44888 CTTCCTTATATAGTGAAGAAGAGTTCTTCTATAGAAGAGGGG-----GGTGGGACAAAGG-- | 44940 |
| QY | 3416 GTTGGGTCTCTGCACGACGCTTGCACAACTTCTTAAATCTCCAGGTGTGCTGCC | 3475 |
| DB | 44941 --TGCCTCTTTCGCAGCAGCTTGCACAACTTCCCTAGATCTCCAGGTGACAGCTGCC | 44998 |
| QY | 3476 TCTTCACAGCAGGTAAAGCAATTTGGTGGGGACATGGTGACCAAGGTGGTGGAGGG | 3535 |
| DB | 44999 TCCCAAGACAGTAAAGCAACCTGGTGGGAACATGAGACCTTAAAGTGGTGGAGGG | 45058 |
| QY | 3536 GACAGGAGTCT--TGCCTCTCTCTGCAGCCT--GTCTTCTTCTTACACCTTGGTAAAT | 3593 |
| DB | 45058 GACGGGGTCTCTGCTTTTCTTGGGGCACTTGGGGCTCTCTGAGACACTTGGTACGGG | 45118 |
| QY | 3594 TTTGGGGGTGAGTAAAGTGTCTCT--GAACCTGAAAGAAGCAAGAGCC--AGCAGGCTG | 3650 |
| DB | 45119 TCGAAGGGGAGGCTTGACACTCTGTGATAGCGTGAAGAAGTGGGAGGCCAATGTGTCTG | 45178 |
| QY | 3651 TCTTGGGCTTCAATGAAGAAGTTCACAGACCCCTTTCTCTGTA--GTACACTTGGT | 3708 |
| DB | 45179 CCTTGGGCTCTCAGGGAGAGACTTTTTCAGGGCCCTGCCCCAGACATAGCTTCTCAGTC | 45238 |
| QY | 3709 TCATCTGTGTAGATTCCCTGGGACCAAGGTGCT-----CTTGGAGCTGAGATT | 3757 |
| DB | 45239 TCTTACCTGTGTGCTTACCCCATCTTGGGTTGTACACATCCCATTTGCCCTGGAACT | 45298 |
| QY | 3758 TCTACATTTAAATCAGGACACTTCTGAG-----ACTTGAATCCGCTGCTGT | 3805 |
| DB | 45359 ACTTCCCGAGCCCCCTCCCAACACTCATCTCTTGGCCGCGCTTCTTCTTGAAGT | 45418 |
| QY | 3863 GG---TTTCTTGTGTACATTTCCCTGACACTCTCTGGAGAGTGTACTTGTGACAT | 3919 |
| DB | 45419 GTTTTCTTCTTGTGTACACATTTCTCTGACACTCTCTGGAGACTCTCGGTTCTGCTCT | 45478 |
| QY | 3920 GTATCTGGGATGTAACTGCAGCCACGAGA 3951 | |

Db 45479 GGCTCTAGCATATGAGCAACACCCACGACA 45510

RESULT 10
AC021016 191754 bp DNA linear PRI 09-MAY-2001
LOCUS AC021016
DEFINITION Homo sapiens BAC clone RP11-378A13 from 2, complete sequence.
ACCESSION AC021016
VERSION AC021016.4 GI:11120952
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 191754)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9847074
PUBMED 9847074
2 (bases 1 to 191754)
Cordes,M., Maupin,R., Hawkins,M. and Boyer,E.
The sequence of Homo sapiens BAC clone RP11-378A13
Unpublished
3 (bases 1 to 191754)
Waterston,R.H.
Direct Submission
Submitted (12-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 191754)
Waterston,R.
Direct Submission
Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 191754)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gi:7650969.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0378A13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frenken,E., Tateo,M., Catanese,J.J. and de Jong,P.J. (1998) An improved

approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP111-146N10; the clone sequenced to the right is RP11-36003. Actual start of this clone is at base position 1 of RP11-378A13; actual end is at base position 191754 of RP11-378A13.

FEATURES
source The sequence RP11-378A13 from 156426 to 156433 is derived from a single ml3 subclone. Size of the region was confirmed by PCR from BAC DNA.

1..191754 location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-378A13"
/clone_lib="RPc1-11"
1..224
/note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"
1..224
/note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"
1..224
/note="similar to EST A157236 (NID:g4310105) tk01d06.x1"
1..216
/note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"
1..216
/note="similar to EST AW020425 (NID:g5873955) df09e11.y1"
2..223
/note="similar to EST AA175052 (NID:g1756173) ms82f02.r1"
2..223
/note="similar to EST AV079022 (NID:g5198850) "
2..223
/note="similar to EST AW538472 (NID:g7180889) "
2..49
/note="similar to EST AV330504 (NID:g6370956) "
7..224
/note="similar to EST A1454294 (NID:g4292719) "
7..224
/note="similar to EST AW433991 (NID:g6965298) "
12..223
/note="similar to EST AW059787 (NID:g6652109) "
20..223
/note="similar to EST A1709077 (NID:g4998853) as86d11.x1"
60..220
/note="similar to EST AA870364 (NID:g2965809) vg44e05.r1"
67..223
/note="similar to EST AV330504 (NID:g6370956) "
447..660
/rpt_family="Alu"
1182..1425
/rpt_family="L1"
1442..1583
/rpt_family="MER1-type"
1620..1928
/rpt_family="Alu"
1948..2347
/note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"
2162..2293
/rpt_family="Alu"
2297..2587
/rpt_family="Alu"
2597..2804
/rpt_family="Alu"
2606..2923
/note="similar to EST AW351985 (NID:g6849698) "
2630..3135
/note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"

| | | | | | |
|----|-----------------------|--|-------------------------|------------------------------|-------------|
| | misc_feature | 2804..3242 | /note="similar to | EST AI916611 (NID:g5636456) | wa28f08.x1" |
| | misc_feature | 2976..3245 | /note="similar to | EST AI401669 (NID:g4244756) | th24e10.x1" |
| | repeat_region | 3097..3284 | /rpl_family="MIR" | | |
| | misc_feature | 3546..3781 | /note="similar to | EST AA102341 (NID:g1647237) | zm26h05.r1" |
| | repeat_region | 3848..4127 | /rpl_family="Alu" | | |
| | repeat_region | 4187..4469 | /rpl_family="Alu" | | |
| | repeat_region | 4474..4690 | /rpl_family="Alu" | | |
| | misc_feature | 4723..5241 | /note="similar to | EST AW444974 (NID:g6986736)" | |
| | misc_feature | 4931..5413 | /note="similar to | EST AA100813 (NID:g1647248) | zm26h05.s1" |
| | repeat_region | 6006..6137 | /rpl_family="MIR" | | |
| | repeat_region | 6243..6394 | /rpl_family="MIR" | | |
| | misc_feature | 6967..7092 | /note="similar to | EST AW461516 (NID:g7031684)" | |
| | misc_feature | 7935..8054 | /note="similar to | EST AW461516 (NID:g7031684)" | |
| | repeat_region | 8294..8433 | /rpl_family="MERL-type" | | |
| | misc_feature | 8553..8604 | /note="similar to | EST AI391439 (NID:g4217443) | tf96e02.x1" |
| | misc_feature | 8556..8749 | /note="similar to | EST AW461516 (NID:g7031684)" | |
| | misc_feature | 9362..9733 | /note="similar to | EST AI391439 (NID:g4217443) | tf96e02.x1" |
| | misc_feature | 9476..9723 | /note="similar to | EST AA595679 (NID:g2411029) | nk66f01.s1" |
| | misc_feature | 9997..10265 | /note="similar to | EST AI634626 (NID:g4685956) | tz30g01.x1" |
| | misc_feature | 10003..10482 | /note="similar to | EST AI828100 (NID:g5448771) | wk31g06.x1" |
| | misc_feature | 10003..10482 | /note="similar to | EST AI890574 (NID:g5595738) | wm87e04.x1" |
| | misc_feature | 10003..10482 | /note="similar to | EST AI978960 (NID:g5803990) | wr68e06.x1" |
| | misc_feature | 10003..10482 | /note="similar to | EST AL047914 (NID:g4728747)" | |
| | misc_feature | 10003..10482 | /note="similar to | EST AW027349 (NID:g5886105) | wf73e09.x1" |
| | misc_feature | 10003..10457 | /note="similar to | EST AI972200 (NID:g5769026) | wr63h01.x1" |
| | misc_feature | 10003..10421 | /note="similar to | EST W48859 (NID:g133688) | zc42c08.s1" |
| | misc_feature | 10003..10330 | /note="similar to | EST AW074215 (NID:g6029213) | xb09d12.x1" |
| | misc_feature | 10003..10249 | /note="similar to | EST AA143563 (NID:g1712935) | zo32d10.s1" |
| | misc_feature | 10003..10247 | /note="similar to | EST AI583619 (NID:g4569516) | ts16d09.x1" |
| | misc_feature | 10003..10240 | | | |
| | Query Match | 2.1% | Score 188; DB 9; | Length 191754; | |
| | Best Local Similarity | 67.4% | Pred. No. 3.6e-43; | | |
| | Matches 507; | Conservative 0; | Mismatches 190; | Indels 55; | Gaps 15; |
| OY | 3328 | AGAGCCAGGCGGGGCGACGCGACGCTGTACAGCTGTCTGAGCTT | 3297 | | |
| | | | | | |
| Db | 164808 | AAAGCCAAGGGGGGGGACACA--TCAGCTCTCATGTCTGAGGCTG--TGGAGGGCTGTGGCT | 164864 | | |
| OY | 3298 | CCCAGGAGCTTGG--GCACGTACTTCCACACCCGCCATTCCTGCTGGGGGCGCAT | 3355 | | |
| | | | | | |
| Db | 164865 | CCCAAGGACCTTGGACGACCTGCTCCCAACCCCTCTCAC-----CTCTCTGGGGGCGCAT | 164919 | | |
| OY | 3356 | CTTCCCTATATGTGAAGAAGTCTCTGGGGGGGGGGGTGCTGAGCAAGAAGTC | 3415 | | |

Db 164920 GTTCCCTATATAGTGGAAGCGATTCTTATGAGAGAGGGG-----GGTGGGCAAAAG-- 164972

Qy 3416 GTTGGCTCCCTGCAGCCAGCTTGGCACAACTTTCTAAAGATCTCCAGAGTGGTGGCTGC 3475

Db 164973 --TGCCTCTTCTGCAGCCAGCTTGGCACAACTTCCCTAGATCTCCAGTGGCACTGCC 165030

Qy 3476 TCTTCCAGACAGTAAAGCAATTTGGGTGGGACACATGGTGTGACCAAGTGGTGGAGG 3535

Db 165031 TCCCCAGACAGTAAAGCAACCTGGTGGGACACATGAGACCTTAACTTGGTGGAGGG 165090

Qy 3536 GACAGAGTCTT-TGCTTCTCTTGGCAGCTT-GTGCTTTCTATGACACTTGGTAACT 3593

Db 165091 GACGGGGTCTCTGCTTTCTCGGGCAACTTGGGGCTCTCTGAGACACTTGTGTGAGGGC 165150

Qy 3594 TTGGGGGTAGTAAAGTGTCTCT-CAAACTCTGAAAGAAAGCAAGAGCC--AGCAGGGCG 3650

Db 165151 TCGAAGAGGGAGGCTTGACACTCTGTGATACCTGTAGAGAGTGTGGGGAGCCCACTGTGTCTG 165210

Qy 3651 TCTTGGGCTTCAATGAAGAGAACTTACAGACAGCCCTTCTCTGTAA--GTACACTTGCCT 3708

Db 165211 CCTTGAGGCTCCAGGGAGAGACTTTTTCGAGGGCCCTGCCCCACATATGCTCTCAGTGC 165270

Qy 3709 TCATCTGTGTAGATTCCCTGGAGCAAGTGGCT-----CTGGAGCTCAGATT 3757

Db 165271 TCCCTACGTGTGGCTAGCCCACTCCCACTTTGGGGTGTACACATCCCAATTTGCCCTGGAAGCT 165330

Qy 3758 TCTACAAATTAAATGAGGACAGTCTCTGAG-----ACTTGGAGTCCGCTGCTGT 3805

Db 165331 TCTAAGAGTTAAACACAGGACAGTGTGTGGGAGACCGGAGCTGTGTGTCTTCCCTACCTGC 165390

Qy 3806 ATTTC----TACTTCTCTGTGCTGCTCTCATTTCTGTATCATGTATACATCTGAAT 3862

Db 165391 ACTCCCGCAGGCCCCCTCCACACACTATCTCTTGGCCGTGCTTCTTATCTGAAGT 165450

Qy 3863 GG---TTTCTTTGTGTACCAATTCCTGTGACACTCCTGGAGTGTGTATCTTGGACAT 3919

Db 165451 GTTTTCTTCTTGTGTACCAATTCCTGTGACACTCCTGGAGTGTGTATCTTGGACAT 165510

Qy 3920 GTATCTGGAGTAAAGTGCAGCCACACAGA 3951

Db 165511 GGCTCTAGCATATAGACACAGCCACACAGA 165542

RESULT 11

AC115965

LOCUS

DEFINITION Mus musculus clone RP24-69114, WORKING DRAFT SEQUENCE, 8 ordered pieces.

ACCESSION AC115965

VERSION AC115965.2 GI:21431261

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 220909)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Unpublished

JOURNAL 2 (bases 1 to 220909)

REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barra,N., Bastien,V., Bloom,T., Boguski,M., Brown,A., Brown,A., Camarata,J., Campiolo,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,M., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Haas,B., Horton,L., Hulme,W., Ilijev,I., Johnson,R., Jones,C., Kanai,A., Karatas,A., Kellis,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Linblad-Ton,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McArthur,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 220909)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N., Anderson,S., Barina,N., Bastien,V., Bloom,T., Bonuskavsky,I., Boukhalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Labroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Marquis,N., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicoli,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strahaus,N., Subramanian,A., Talamas,J., Teste,S., Theodore,J., Topham,K., Travers,N., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 17, 2002 this sequence version replaced gi:19683644.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L24961
Center clone name: 69.I.14

----- Summary Statistics -----
Sequencing vector: Plasmid; n/a: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 219013 bases at least Q40
Consensus quality: 219547 bases at least Q30
Consensus quality: 219760 bases at least Q20
Insert size: 212000; agarose-fp
Insert size: 220209; sum-of-ctrls
Quality coverage: 9.3 in Q20 bases; agarose-fp
Quality coverage: 9.0 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. -----
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 2262: contig of 2262 bp in length

FEATURES
source
1..220909
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-69114"
/clone_11b="RP24 Male Mouse BAC"

misc_feature
1..2262
/note="assembly-fragment"

misc_feature
2363..5650
/note="assembly-fragment"

misc_feature
5751..14729
/note="assembly-fragment"

misc_feature
14830..29889
/note="assembly-fragment"

misc_feature
29990..51480
/note="assembly-fragment"

misc_feature
51581..87582
/note="assembly-fragment"

misc_feature
87683..146358
/note="assembly-fragment"

misc_feature
146459..220909
/note="assembly-fragment"

BASE COUNT 60812 a 48551 c 48344 g 62502 t 700 others

ORIGIN

Query Match 2.1% Score 184.8; DB 2: Length 220909;
Best Local Similarity 68.3% Pred. No. 3.2e-42;
Matches 351; Conservative 0; Mismatches 142; Indels 21; Gaps 6;

1839 AAAAGTATATGTTGGAGAGAGATGCTCCAGAGACTTCTGCTCTTC 1898
118835 AAGAAATCTTTTGGAGTTGGAGATGCTCAGCATTTAGAAACA----CTGCTCTTC 118890
1899 AGAGGACCTGATTCAGTCCAGGACTCATATGTTGGCTCAGACGACCATCTGTAATCCA 1958
118891 CAAGGACGAGTTTCAACCTCCAGCATCTACAGAGTGGTCATTAATCATCTGCAACTCAA 118950
1959 GTTCCAGAGGCTTCCACACCTCTTCTGCGCTGCACAGCAGACACAT -ACATAGTTCAC 2016
118951 GTTCCACATATATCCAAATCTGTTTGGCTTCTGAGGAGCAGACATATGATGTCAC 119010
2017 AGACATCATGACGAGCAAAACACCATCATATTAATTAAGAACTTAAGAGT 2076
119011 AGATCACGTTGACGCAAAACACCATCATCTGTAAGAAATTAATTTGATGTTAACCC 119070
2077 GCATGTTGTTGTTAACAATTTGCTTACACATGCTGATTTGAAGACATGTACAGGCACACA 2136
119071 G-TGGCCTCAGTCACTGATATGATTAATTAAGTGAAGAAACATTTGATGTGA--GAATGTT 119127
2137 CTGAAGAGGATCTGGGCTGAGAGATGCTCAGCGTTTAAGAGCCTACGCTCTTC 2196
119128 TTTAAGAACTCACTGGGCTGAGAGATGCTCAGGTTTAAGAGTCCCATGCTCTTC 119187
2197 CGAAGGAGGCTGATGATTAATTCATGACACCATGCTGCTCAGCAACATCATCA 2256
119188 C-----AAAGTCCGAGTTTAAATCTGACACCATGTTGCTCAGCAACATCATCA 119243
2257 TGAGATCTGACACCTCTTCTGCTGATCTGAAGACAGCTGACAGCTACAGTACTTA 2316
119243

Db 119244 TGAATCTGATGCCCTCTCTGAGTGTCTGAAGAC-----AGCTACGTACTTA 119295
 QY 2317 GATATACATATAATAATCTTTTAAAAAA 2350
 Db 119296 CATATATAATAATAATCTTTAAAAAA 119329

RESULT 12
 AC102777/c
 LOCUS AC102777 184591 bp DNA linear HTG 21-AUG-2002
 DEFINITION Mus musculus clone RP23-5K7, WORKING DRAFT SEQUENCE, 25 unordered
 pieces.
 AC102777
 AC102777.2 GI:22381791
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 184591)
 Birren,B., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-5K7
 Unpublished
 2 (bases 1 to 184591)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Lamaze,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarty,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
 Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Roselt,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 184591)
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
 Batra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cooke,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karats,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarty,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zemek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:17060887.
 All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19649
 Center clone name: 5_K-7
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator big dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 176248 bases at least Q40
 Consensus quality: 180182 bases at least Q30
 Consensus quality: 181502 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 182191; sum-of-coverage
 Quality coverage: 5.8 in Q20 bases; agarose-fp
 Quality coverage: 6.0 in Q20 bases; sum-of-coverage

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 806: contig of 806 bp in length
 * 807 906: gap of 100 bp
 * 907 1108: contig of 202 bp in length
 * 1109 1208: gap of 100 bp
 * 1209 1890: contig of 682 bp in length
 * 1891 1990: gap of 100 bp
 * 1991 2684: contig of 694 bp in length
 * 2685 2784: gap of 100 bp
 * 2785 3648: contig of 864 bp in length
 * 3649 3748: gap of 100 bp
 * 3749 5130: contig of 1382 bp in length
 * 5131 5230: gap of 100 bp
 * 5231 6203: contig of 973 bp in length
 * 6204 6303: contig of 100 bp
 * 6304 7763: contig of 1460 bp in length
 * 7764 7863: gap of 100 bp
 * 7864 8416: contig of 553 bp in length
 * 8417 8516: gap of 100 bp
 * 8517 11254: contig of 2738 bp in length
 * 11255 11354: gap of 100 bp
 * 11355 12417: contig of 2063 bp in length
 * 12418 13517: gap of 100 bp
 * 13518 15594: contig of 2077 bp in length
 * 15595 15694: gap of 100 bp
 * 15695 19567: contig of 3873 bp in length
 * 19568 19667: gap of 100 bp
 * 19668 22294: contig of 2627 bp in length
 * 22295 22394: gap of 100 bp
 * 22395 26269: contig of 3875 bp in length
 * 26270 26369: gap of 100 bp
 * 26370 29166: contig of 2797 bp in length
 * 29167 29266: gap of 100 bp
 * 29267 34448: contig of 5182 bp in length
 * 34449 34548: gap of 100 bp
 * 34549 43348: contig of 8800 bp in length
 * 43349 43448: gap of 100 bp
 * 43449 51979: contig of 8531 bp in length
 * 51980 52079: gap of 100 bp
 * 52080 67532: contig of 15453 bp in length
 * 67533 67632: gap of 100 bp
 * 67633 79233: contig of 11601 bp in length
 * 79234 79333: gap of 100 bp
 * 79334 96823: contig of 17490 bp in length

| | | | |
|---|--------|-------------------|--------------------|
| * | 96624 | 96923: gap of | 100 bp |
| * | 96924 | 122005: contig of | 25082 bp in length |
| * | 122006 | 122105: gap of | 100 bp |
| * | 122106 | 161850: contig of | 39745 bp in length |
| * | 161851 | 161950: gap of | 100 bp |
| * | 161951 | 184551: contig of | 22641 bp in length |

```
FEATURES      Location/Qualifiers
source        1. .184591
```

| | |
|----------------|---------------------------------------|
| misc_feature | /clone.lib="RPCT-23 Female Mouse BAC" |
| 1..806 | |
| misc_feature | /note="assembly_fragment" |
| 907..1108 | |
| misc_feature | /note="assembly_fragment" |
| 1209..1890 | |
| misc_feature | /note="assembly_fragment" |
| 1991..2684 | |
| misc_feature | /note="assembly_fragment" |
| 2785..3648 | |
| misc_feature | /note="assembly_fragment" |
| 3749..5130 | |
| misc_feature | /note="assembly_fragment" |
| 5331..6203 | |
| misc_feature | /note="assembly_fragment" |
| 6304..7763 | |
| misc_feature | /note="assembly_fragment" |
| 7864..8416 | |
| misc_feature | /note="assembly_fragment" |
| 8517..11254 | |
| misc_feature | /note="assembly_fragment" |
| 11355..13417 | |
| misc_feature | /note="assembly_fragment" |
| 13518..15594 | |
| misc_feature | /note="assembly_fragment" |
| 15695..19567 | |
| misc_feature | /note="assembly_fragment" |
| 19668..22294 | |
| misc_feature | /note="assembly_fragment" |
| 22395..26269 | |
| misc_feature | /note="assembly_fragment" |
| 26370..29166 | |
| misc_feature | /note="assembly_fragment" |
| 29267..34448 | |
| misc_feature | /note="assembly_fragment" |
| 34549..43348 | |
| misc_feature | /note="assembly_fragment" |
| 43449..51979 | |
| misc_feature | /note="assembly_fragment" |
| 52080..67532 | |
| misc_feature | /note="assembly_fragment" |
| 67633..79233 | |
| misc_feature | /note="assembly_fragment" |
| 79334..96823 | |
| misc_feature | /note="assembly_fragment" |
| 96924..112005 | |
| misc_feature | /note="assembly_fragment" |
| 122106..161850 | |
| misc_feature | /note="assembly_fragment" |
| 161951..184591 | |
| misc_feature | /note="assembly_fragment" |
| clone_end:17 | |

| | | | | |
|-----------------------|-----------------|--------------------|------------|----------------|
| Query Match | 2.0%; | Score 184; | DB 2; | Length 184591; |
| Best Local Similarity | 69.5%; | Pred. No. 5.3e-42; | | |
| Matches 346; | Conservative 0; | Mismatches 125; | Indels 27; | Gaps 6 |

OY 1847 TATGTTGGAGGAGAAATGGCTCAGCTTCACAGAGACTTGTGCTCTTGACAGAGACC 1906
| | | | | | | | | | | | | | | | | | | | | |
Db 89227 TGTCATTGCTGTGTTTAAGAAGCTCAGTGGTCAAGAGCACTGGCTGCTTCACAGAGAAC 89168

OY 1307 TAGATTTCAGTGTCCACAGACTCATATATGTTGGCTCACAGGCATCTGTAATATCCAGTTCCAGA 1966

| | | | |
|----|-------|---|-------|
| Db | 89167 | AGGGCTCAGTTCCTCAATAACCAAGATGATGGCTCCACAACTATTGTAACTCCAGTCCGAGG | 89108 |
| Qy | 1967 | GAGGTTCACACACCCTCTTGCGCTCCACAGGACCACACATACA-TAGTACAGACATACA | 2025 |
| Db | 89107 | GCATGTGGGTGGCCCTCTTCTGGGCTTCACAGGAAATTTCATTACAAAGTACACAGACATCA | 89048 |
| Qy | 2026 | TGCAGGCAAAAACCCCATACACACATAAATAATAATAGCAAACTTAAAGGTGCATGTGT | 2085 |
| Db | 89047 | TGC-----AAACACATCACATACATAAATCAATTCAATTAAATTAAATACACATAAAT | 88993 |
| Qy | 2086 | GGTAAACATTGTGCTTACACATGCTGATTTGAAGACATGTACACGCACACACTGAAGAG | 2145 |
| Db | 88992 | GAT-AACAAAAGCTGCTCCAGAAATTGAGTAAATAATAATTAATGCTTAAATTTCTG----- | 88940 |
| Qy | 2146 | GATCTGGGGGCTGGAGAGATGGCTACGGCGTTAAGAGCACTGACGTCTTCCGAAGAG | 2205 |
| Db | 88939 | --GTAGGGGCTGGAGATGATGGCTCAGTGATTAAGAGCACTGACGTCTTCTG---TGAG | 88886 |
| Qy | 2206 | GTCCTGAGTTCAAAATCCATAGCAACACACATGGTGGCTCAACACCATCCATTAATGAGANTG | 2265 |
| Db | 88885 | GTCCTGAGTTCAAAATCCAGCAACACACATGGTGGCTCAACACCATTAATGAGANTG | 88826 |
| Qy | 2266 | ACACCCTCTTCTGGTGACATCTGGAAGACAGCTGGCAGAGCTTAAGTGTACTTAATATACTA | 2325 |
| Db | 88825 | ACTCCCTCTTCTGGGTGTGTGAAGAC-----AGCTACAGTGTACTACATATATA | 88774 |
| Qy | 2326 | ATAAATAAATCTTTTTT 2343 | |
| Db | 88773 | ATAAATCTTTTTTTTTT 88756 | |

Db 88773 AATAAATCTTTT TTTT 88756

| | | | | |
|------------|--|-----------|-----|-----------------|
| RESULT 13 | | | | |
| AF100956/c | | | | |
| LOCUS | AF100956 | 273800 bp | DNA | linear |
| DEFINITION | Mus musculus major histocompatibility locus class II region; | | | |
| | | | | ROD 03-NOV-1998 |

Fas-binding protein Daxx (DAXX) gene, partial cds, Bln1 (RING1), tapasin (tapasin), RALGDS-like factor (RLF), KE2 (KE2), BING4 (RING4), beta1, 3-galactosyl transferase (beta1,3-galactosyl transferase), ribosomal protein subunit S18 (RP518), Secm21 (Secm21), H2A1(b) (H2-K1(b)), RING1 (RING1), KE6a (KE6a), KE4 (KE4), Rxbbeta (Rxbbeta), collagen alpha-2 (X1) (COL1A2), H2-O alpha (H2-O-alpha), RING3 (RING3), H2-M alpha (H2-M alpha), H2-M beta 2 (H2-M beta2), and H2-M beta1 (H2-M beta1) genes, complete cds; and Irfp 2 gene, partial cds.

| | |
|-----------|------------|
| ACCESSION | AF100956 |
| VERSION | AF100956.1 |
| | GI:3811374 |

KEYWORDS
SOURCE
ORGANISM

REFERENCE
Eukaryota: Metazoa: Chordata: Crustacea: Euteleostomi:
Mammalia: Eutheria, Rodentia, Sciurognathi, Muridae, Mus.
1 (bases 1 to 273800)

| TITLE | AUTHORS |
|---|---|
| Sequence of the mouse major histocompatibility locus class II | Rowen, L., Qin, S., Medan, A., Loretz, C., James, R., Dors, M., Mix, L., Hall, J., Lasky, S. and Hood, L. |

region
unpublished
JOURNAL
REFERENCE
2 (bases 1 to 273800)

AUTHORS Rowen, L.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1998) Department of Molecular Biotechnology, Box 357730 University of Washington, Seattle, WA 98195, USA
COMMENT This sequence overlaps the entry in GenBank accession Number

AF022765 by 22022 bases. Sequencing methodology: high redundancy shotgun, using M13 templates. Interspersed Repeats were identified with RepeatMasker (available from <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

```

FEATURES
  source
    location/Qualifiers
      1. 273800
        /organism="Mus musculus"
        /strain="129svj"
        /db_xref="taxon:10090"

```

```

source
    /chromosome="17"
    1. .139758
    /organism="Mus musculus"
    /strain="129svJ"
    /db_xref="taxon:10090"
    /clone="BAC 445D10"
    /clone.lib="Genome Systems"
    139672. .140387
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="BAC 346M2"
    /clone.lib="Research Genetics"
    /note="A PCR product made from BAC 346M2 was sequenced to
    fill the gap between BACs 445D10 and 36B7"
    140323. .273800
    /organism="Mus musculus"
    /strain="129svJ"
    /db_xref="taxon:10090"
    /clone="BAC 36B7"
    /clone.lib="Genome Systems"
    <1. .2000
    /gene="DAXX"
    join(<146. .208,325. .544,1825. .2000)
    /product="Fas-binding protein Daxx"
    join(<146. .208,325. .544,1825. .1884)
    /gene="DAXX"
    /note="Intron-exon boundaries defined in relation to cDNA
    sequence in GenBank Accession Number AF006040"
    /codon_start=2
    /product="Fas-binding protein Daxx"
    /protein_id="AAC69891.1"
    /db_xref="GI:3811375"
    /translation="PPSKRFRKKQGLGNSYIKPEPAQDSGNTSVQPMPS
    PLASVAVSDSTRVDSPSHELVTSLCSPSPSLLOTPQASLRQCIYKTSVATQCD
    PEELIVLSDSD"
    724. .858
    /rpt_family="MMB1"
    complement(1257. .1297)
    /rpt_family="(CA)n"
    1303. .1421
    /rpt_family="MMB1F"
    complement(1426. .1630)
    /rpt_family="B3"
    complement(1646. .1792)
    /rpt_family="MMB1F"
    complement(2728. .2788)
    /rpt_family="GC-rich"
    3168. .3214
    /rpt_family="MIR"
    <3365. .>5281
    /gene="BIN1"
    /note="zn finger; mouse orthologue to BIN1, described in
    GenBank Accession Number z97183"
    <3365. .>5281
    /product="Bing1"
    /gene="BIN1"
    /note="this ort was predicted by Genscan and confirmed by
    ESTs AA538332 and AA797063"
    /codon_start=1
    /product="Bing1"
    /protein_id="AAC69892.1"
    /db_xref="GI:3811376"
    /translation="MEPSALSPSGATLPLPLAPPLPAPAAVHVHPEEVSAL
    ESLNOORLOGLDYSIRVOGREFAHRAVAASSPYHDVLLKGMSTLSPTMDF
    GAFETVLASATYGRISMAADIVNLTGVSYLQMHYDKTELLREBSAATTTTV
    TAAAPSVPKASVPSGNGTVAPTVGSVSHSSRSSENQSPSSNFPSPRESTDF
    SSTGDAFVAVASAGGNRRDGPVPFAPVVSAGTSGRLLEADLDCDDGAGV
    AFGAGILRSNCMPASVAPQKHVVYKQARNCPPASISLHODPDEDEEDDLVTCE
    DDEDEMGSGGSPAGGEATLISIDVRTLPEADKGEQVNCESNDGCPYEGCG
    AGAGLDPRGPTPSSYALTHPRPLPLDVPGNQILVPPSSSQAPQGPENTAEHGA

```

```

gene
    /gene="tapasin"
    join(6121. .6206,6353. .6532,6731. .6991,11888. .12286,
    12486. .12827,13108. .13184,14618. .14754)
    /gene="tapasin"
    /product="tapasin"
    join(6170. .6206,6353. .6532,6731. .6991,11888. .12286,
    12486. .12827,13108. .13184,14618. .14629)
    /gene="tapasin"
    /note="intron-exon boundaries defined in relation to cDNA
    in GenBank Accession Numbers AF043943 and EST AA565898;
    EST W75528 contains an alternatively spliced form of this
    transcript"
    /codon_start=1
    /product="tapasin"
    /protein_id="AAC69893.1"
    /db_xref="GI:3811377"
    /translation="MKPLLLVAVAGLATVSVVSAGPEALIEGMYEDAGGGLSK
    PATLLLRHGRPGPRPDLDKLYEKVDPAGMLAARRYPAGASAPHCMSRTIP
    PASAKWARSLSPEQNCPRALDMLVSVSTLSLSLRPOPELREPVITMATV
    VLTVLTNHPARVOLGDAVLDLRPAVPSALESPSLDAGPPGLEMRORHGKCH
    LLIAATRGILAGMRVADKATAPFAMDDPEWCPWCTGTMPLAVKRSQESGLATV
    HLPYIQGVSLIELTVHKAIRYSLTIPAPVYVAAAGAPPELLCLASHFPDAGLEKWE
    LRGGCGSSRVEKGTWLTIRHSDGVSOSGHLQLPVPAKOHGVYVGVYHSSL
    PASGSAVLTLEVAAYWTIEVSEKATASLTIPRNSKKSQ"
    complement(8176. .8383)
    /rpt_family="B3"
    845. .8670
    /rpt_family="B3"
    complement(9036. .9243)
    /rpt_family="MMB2"
    complement(9853. .9997)
    /rpt_family="MMB1"
    complement(10003. .10198)
    /rpt_family="MMB2"
    10003. .10014
    /gene="tapasin"
    /note="number of Ts unclear"
    10369. .10460
    /rpt_family="(GA)n"
    10454. .10474
    /gene="tapasin"
    /note="low quality data"
    complement(10836. .10943)
    /rpt_family="PBID7"
    10973. .11020
    /rpt_family="MUSIDS"
    11215. .11328
    /rpt_family="MMB1F"
    11329. .11427
    /rpt_family="(GA)n"
    complement(11428. .11459)
    /rpt_family="(CA)n"
    11467. .11637
    /rpt_family="B3"
    complement(13691. .13830)
    /rpt_family="MMB1"
    complement(13859. .13993)
    /rpt_family="MWSOR2"
    complement(14790. .14932)
    /rpt_family="MMB1F"
    complement(14947. .15087)
    /rpt_family="MMB1"
    complement(15009. .15110)
    /rpt_family="MTE2"
    15133. .15243
    /rpt_family="MMB1"
    complement(15387. .15490)
    /rpt_family="UR1"

```

```

gene
16444..24873
/gene="RLF"
/note="described as a Ras-associating protein in GenBank"
Accession Number U54639"
Join(16444..16522,16723..16919,18204..18287,18518..18696,
18799..18849,18932..19229,19487..19738,19829..19932,
20019..20103,20190..20258,20375..20449,21453..21485,
21612..21733,21838..21933,22023..22134,22218..22511,
23341..23455,23543..23754)
/gene="RLF"
/product="RalGDS-like factor"
Join(16764..16919,18204..18287,18518..18696,18799..18849,
18932..19229,19487..19738,19829..19932,20019..20103,
20190..20258,20375..20449,21453..21485,21612..21733,
21838..21933,22023..22134,22218..22511,23341..23455,
23543..23754)

Query Match      2.08; Score 183.6; DB 10; Length 273800;
Best Local Similarity 62.5%; Pred. No. 7,5e-42;
Matches 339; Conservative 0; Mismatches 194; Indels 9; Gaps 3;

QY 1853 GCGAGGAGAGAGAGCTGACGCTTCAGAGACACTTGTGCTTCAGAGAGAGCTAGATT 1912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107701 GGGCTGGAAGAGATGGCCAGCGGCTTAGAGACACTGGCCTTTGTCAGAGAGAA--TCCGTT 107644

QY 1913 CAGTTCCAGAGACTCATATGATGGCTCAGAGCCATCTGTAATTCAGTTCAGAGAGGCTTC 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107643 TGATTCCTAGATCCACAGAGTACTTCCAAACCATCTGTAACTCAGTGCAGAGCGCTTG 107584

QY 1973 CACACCTCTTCTTGCTCCAGAGGACCATCATATGATACAGACATATCATGACAGGC 2032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107583 GATGCCCTTTTGTACCTCTACACTCTCCAGGAACAGCTGGTACACAGTGGTGCAGGC 107524

QY 2033 AAAACACCCTATACACATTAATTAAGAAACCTTAAAGAGTGCATGTGTTGCTTAAC 2092
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107523 AAAACACCAACACACAGGATTTAAAA---TAAGCTAGTTGAGTGGGTGATTTAAAGGT 107467

QY 2093 ATTGTGCTTACATGCTGATTTGAAGACATGTACACGACACACACTGAAAGAGGATCTGG 2152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107466 GAGGCGAGGCGGCTCTCTGCAAGACTAAGTAGTAAGAACCTGCTTTAAAAAACACCAAG 107407

QY 2153 GCGTGGAGAGATGGCTCAGCGGTTAAAGACACTGCTCTTCCGAAGAGAGTCTCTGA 2212
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107406 GGTGGAGAGAGTGGCTAGTGGTTAAAGACAGGACGCTGCTTCCAG---AGGCTTGA 107351

QY 2213 GTTCAATTCCTAGACACACATGCTGGCTCACAACCATTAATGATGATCTGACACCT 2272
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107350 GTTCAATTCCTAGACACACATGCTGGCTCACAACCATCTGTAGTGGATCTGTATGCCCT 107291

QY 2273 CTTCCTGTGATCTGAAGACAGCTGACAGCTACAGTGTACTTGAATATTAATAATA 2332
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107290 TTTCCTGTGATCTGAAGACAGCTGACAGTGTACTTGAATATTAATAATA 107231

QY 2333 AATCTTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAAGAGATTATGACAGCT 2332
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107230 AAAAAAGATAAACAAAAAACCCCCCAAAAACACAAACAAACACACTTTTAAATTGG 107171

QY 2393 GA 2394
    ||
Db 107170 GA 107169

```

```

RESULT 14
AC130278/c 182695 bp DNA linear HTG 09-AUG-2002
LOCUS Mus musculus clone RP24-393116, WORKING DRAFT SEQUENCE, 36
DEFINITION Unordered pieces.
AC130278
VERSION AC130278.1 GI:22165183
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 182695)
TITLE Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Mus musculus, clone RP24-393116
AUTHORS Unpublished
REFERENCE 2 (bases 1 to 182695)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Chepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karalas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schnpack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfay,S., Theodore,J., Topham,K., Travers,Y., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zalnoun,J.,
Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L25607
Center clone name: 393_1_16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168477 bases at least Q40
Consensus quality: 173980 bases at least Q30
Consensus quality: 176349 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 179195; sum-of-contigs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 7169 7268: gap of 100 bp
* 7269 7773: contig of 505 bp in length
* 7774 7873: gap of 100 bp
* 7874 8549: contig of 676 bp in length
* 8550 8649: gap of 100 bp
* 8650 9264: contig of 615 bp in length
* 9265 9364: gap of 100 bp
* 9365 10058: contig of 694 bp in length
* 10059 10158: gap of 100 bp
* 10159 10902: contig of 744 bp in length
* 10903 11002: gap of 100 bp
* 11003 11256: contig of 254 bp in length
* 11257 11356: gap of 100 bp
* 11357 12233: contig of 877 bp in length
* 12234 12333: gap of 100 bp

```



```

* 12334 13682: contig of 1349 bp in length
* 13683 13782: gap of 100 bp
* 13783 14874: contig of 1092 bp in length
* 14875 14974: gap of 100 bp
* 14975 16211: contig of 1237 bp in length
* 16212 16311: gap of 100 bp
* 16312 17319: contig of 1008 bp in length
* 17320 17419: gap of 100 bp
* 17420 19489: contig of 2070 bp in length
* 19490 19589: gap of 100 bp
* 19590 21253: contig of 1664 bp in length
* 21254 21353: gap of 100 bp
* 21354 22801: contig of 1448 bp in length
* 22802 22901: gap of 100 bp
* 22902 24454: contig of 1553 bp in length
* 24455 24554: gap of 100 bp
* 24555 25711: contig of 1157 bp in length
* 25712 25811: gap of 100 bp
* 25812 28015: contig of 2204 bp in length
* 28016 28115: gap of 100 bp
* 28116 29791: contig of 1676 bp in length
* 29792 29891: gap of 100 bp
* 29892 32114: contig of 2223 bp in length
* 32115 32214: gap of 100 bp
* 32215 32974: contig of 760 bp in length
* 32975 33074: gap of 100 bp
* 33075 33793: contig of 719 bp in length
* 33794 33893: gap of 100 bp
* 33894 37124: contig of 331 bp in length
* 37125 37224: gap of 100 bp
* 37225 40484: contig of 3260 bp in length
* 40485 40584: gap of 100 bp
* 40585 42991: contig of 2407 bp in length
* 42992 43091: gap of 100 bp
* 43092 48293: contig of 5202 bp in length
* 48294 48393: gap of 100 bp
* 48394 53122: contig of 4729 bp in length
* 53123 53222: gap of 100 bp
* 53223 75713: contig of 22493 bp in length
* 75716 75815: gap of 100 bp
* 75816 83666: contig of 7851 bp in length
* 83667 83766: gap of 100 bp
* 83767 92545: contig of 8779 bp in length
* 92546 92645: gap of 100 bp
* 92646 103351: contig of 10706 bp in length
* 103352 103451: gap of 100 bp
* 103452 118198: contig of 14747 bp in length
* 118199 118298: gap of 100 bp
* 118299 134605: contig of 16307 bp in length
* 134606 134705: gap of 100 bp
* 134706 151709: contig of 17004 bp in length
* 151710 151809: gap of 100 bp
* 151810 175524: contig of 23715 bp in length
* 175525 175624: gap of 100 bp
* 175625 182695: contig of 7071 bp in length.

```

FEATURES

Source

```

1. 182695
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="RP24-393i16"
   /clone.lib="RPCI-24 Male Mouse BAC"
   1. 7168
    /note="assembly_fragment"
    /note_end:SP6
    vector_side:left"
    misc_feature
      7269.. 7773
      /note="assembly_fragment"
    misc_feature
      7874.. 8549
      /note="assembly_fragment"
    misc_feature
      8650.. 9264
      /note="assembly_fragment"
    misc_feature
      9365.. 10058
      /note="assembly_fragment"

```

```

misc_feature 10159..10902
             /note="assembly_fragment"
misc_feature 11003..11256
             /note="assembly_fragment"
misc_feature 11357..12233
             /note="assembly_fragment"
misc_feature 12334..13682
             /note="assembly_fragment"
misc_feature 13783..14874
             /note="assembly_fragment"
misc_feature 14975..16211
             /note="assembly_fragment"
misc_feature 16312..17319
             /note="assembly_fragment"
misc_feature 17420..19489
             /note="assembly_fragment"
misc_feature 19590..21253
             /note="assembly_fragment"
misc_feature 21354..22801
             /note="assembly_fragment"
misc_feature 22902..24454
             /note="assembly_fragment"
misc_feature 24555..25711
             /note="assembly_fragment"
misc_feature 25812..28015
             /note="assembly_fragment"
misc_feature 28116..29791
             /note="assembly_fragment"
misc_feature 29892..32114
             /note="assembly_fragment"
misc_feature 32215..32974
             /note="assembly_fragment"
misc_feature 33075..33793
             /note="assembly_fragment"
misc_feature 33894..37124
             /note="assembly_fragment"
misc_feature 37225..40484
             /note="assembly_fragment"
misc_feature 40585..42991
             /note="assembly_fragment"
misc_feature 43092..48293
             /note="assembly_fragment"
misc_feature 48394..53122
             /note="assembly_fragment"

```

Query Match

Best Local Similarity 65.2%; Score 181.8; DB 2; Length 182695;

Matches 359; Conservative 0; Mismatches 137; Indels 55; Gaps 4;

```

QY 1797 ACATATAGTTACATGATGATATATATATAGTAAATTCCTTTAAAGGATATATGTTGGGA 1856
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171436 ACAGTACTTACATATATATATATATATATATATATATATATATATATATATATATAT 171377
QY 1857 GGGACAGATGGGCTCGCTCCAGGACCTGCTGCTGTCGAGGACCTAGATTCACT 1916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171376 TAGAGAGATGATAGTGTGCTCAGAACCTGCTCTATCCAGGACCTTGTTCAAT 171317
QY 1917 TCCAGAGCATATATGTTGCTTCACAGGCATCTGTAATCCAGTTCCAGAGGTTCCACA 1976
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171316 TCCAGACACATACATGTTGCTCAGACGCTGCTGAGCTAGTGTCCAGGGTCCACAA 171257
QY 1977 CCCTCTTGTGCGCTCCACAGCACCACAT----ACATAGTACACAGACATACATCAGGC 2032
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171256 GTCTCTTGTGCGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 171197
QY 2033 AAAACACCCATACACACATTAATAATTAAGAACTTAAGGTCATGTTGTTAAGC 2092
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171196 AAAACACTCATCATATATATAGTAAATAGTAATGTTTAAAGG----- 171154
QY 2093 ATTGTCTTACACATGCTGATTGAAGACATGTACACACCTGAAGAGGATCTGG 2152
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 171153 -----AAAGGAGGAGAGAAAGAGAGACAGAAAAAAGAACAAAG 171116

```

QY 2153 GCGTGAGAGATGGCTCAGCGGTTAAGAGCAGTCACTGCTTCCGAGCAGAGGTCCTGA 2212
||||| ||||||| || ||||||| ||||| |||||||
Db 171115 GCGGTGTAAGATGGCTGAGGTAAGAGCAGTCACTGCTTCC-----AAAGGTCCTGA 171060
QY 2213 GTTCAATCTAGCAGACCAACCATGTGGCTCACAACCATTCATATGAGATCTGACACCTT 2272
||||| ||||||| || ||||||| ||||| ||||||| ||||| |||||||
Db 171059 GTTCAATCTAGCAGTATCAATGCTGCTTCAACCAACCATTAATGAGATCTGACACCTT 171000
QY 2273 CTTCGTGCGATCTGAGACAGCTGACAGCTACAGTCACTTATGATATCTATAATAATA 2332
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 170999 CTTCGTGCGCTCTGAGAGAC-----ACCTACAGCTGCTTATGATATATAATAATAATA 170948
QY 2333 AATCTTTTTT 2343
||||| |||||||
Db 170947 AATCTTTTTT 170937

RESULT 15
AC025964 218502 bp DNA linear HTG 09-MAY-2001
LOCUS AC025964/c
DEFINITION Mus musculus chromosome 11 clone RP23-136C7 map 11, WORKING DRAFT
SEQUENCE 10 unordered pieces.
AC025964
AC025964.3 GI:13259480
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 218502)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguski,M., Borker,D., Brown,A., Burt,D., Buzan,C., Burt,D.,
Campbell,A., Castelle,A., Chao,J., Cho,J., Collins,S.,
Collins,S., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S.,
Dodge,S., Dominko,T., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardy,J., Ginde,S., Glendon,S., Goff,L., Graham,L.,
Grand-Pierre,N., Grant,G., Haggis,B., Heath,L., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karas,A.,
Klein,J., Lacroix,K., Lamazares,R., Landers,T., Lehoucq,J.,
Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherson,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Miyauchi,M., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teste,S., Theodore,J., Tirrell,A., Travers,M., Triggiani,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
COMMENT Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 9, 2001 this sequence version replaced g1:880918.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5729
Center clone name: 136C-7
----- Summary Statistics
Sequencing vector: M13; M77815; 38% of reads
Sequencing vector: Plasmid; n/a; 62% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 215619 bases at least Q40
Consensus quality: 216900 bases at least Q30
Consensus quality: 217250 bases at least Q20
Insert size: 211000; agarose-fp
Insert size: 217602; sum-of-contigs
Quality coverage: 8.2 in Q20 bases; agarose-fp
Quality coverage: 7.9 in Q20 bases.
NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 20205: contig of 20205 bp in length
* 20206 20305: gap of 100 bp
* 20306 25641: contig of 5336 bp in length
* 25642 25741: gap of 100 bp
* 25742 35304: contig of 9563 bp in length
* 35305 35404: gap of 100 bp
* 35405 45241: contig of 9837 bp in length
* 45242 45341: gap of 100 bp
* 45342 62833: contig of 17492 bp in length
* 62834 62933: gap of 100 bp
* 62934 110480: contig of 47547 bp in length
* 110481 110580: gap of 100 bp
* 110581 137309: contig of 26729 bp in length
* 137310 137409: gap of 100 bp
* 137410 166669: contig of 29660 bp in length
* 166670 166769: gap of 100 bp
* 166770 201541: contig of 34772 bp in length
* 201542 201641: gap of 100 bp
* 201642 218502: contig of 16861 bp in length.
*
FEATURES
source
1. 218502
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-136C7"
/clone_lib="RPC1-23 Female Mouse BAC"
1. 20205
/note="assembly-fragment
clone_end:SP6
vector_side:left"
20306. 25641
/note="assembly-fragment"
25742. 35304
/note="assembly-fragment"
35405. 45241
/note="assembly-fragment"
45342. 62833
/note="assembly-fragment"
62934. 110480
/note="assembly-fragment"
110581. 137309
/note="assembly-fragment"
137410. 166669
/note="assembly-fragment"
166770. 201541
/note="assembly-fragment"
201642. 218502
/note="assembly-fragment
clone_end:T7
vector_side:right"
misc_count 57976 a 50523 c 50316 g 58775 t 912 others
ORIGIN
Query Match 2.0%; Score 179.2; DB 2; Length 218502;
Best Local Similarity 61.4%; Pred. No. 1.4e-40;

THIS PAGE BLANK (USPTO)

PN WO200034492-A1.
XX 15-JUN-2000.
XX
PE 09-DEC-1998; 98WO-EP08009.
XX 09-DEC-1998; 98WO-EP08009.
PR (CNRS) CENT NAT RECH SCI.
PA (CURIE) INST CURIE.
XX
PI Pinto D, Robline S, Jaisser F, Louvard D;
XX
DR WPI; 2000-423433/36.
XX
PT Novel nucleotide sequence derived from mouse villin gene for targeted
PT expression of transgenes in immature and differentiated epithelial
PT cells of intestine or urogenital tracts .
XX
PS Claim 3; Fig 6; 54pp; English.
XX
CC The present sequence comprises the murine villin gene regulatory
CC region, the first intron and part of the first two exons. It has been
CC shown that this region directs the expression of the villin gene in the
CC intestine and uro-genital tracts, and thus could be used in a fusion
CC gene to direct expression of exogenous genes in these areas. This could
CC be used, for example, to create a mouse model for colorectal cancer.
XX
SQ Sequence 8995 BP; 2275 A; 2105 C; 2258 G; 2357 T; 0 other:

Query Match 100.0%; Score 8995; DB 21; Length 8995;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 8995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTGCTGCACCAAGACACTGTGTCGCCAGCACTGGGAGTGAGGAGAGAGGTCA 60
DB 1 GATCTGCTGCACCAAGACACTGTGTCGCCAGCACTGGGAGTGAGGAGAGGTCA 60
QY 61 GAACCTTAAGGTCATCCCTGGTTCATAGCAAGGTTTCAGGCACTTCAGCTACATGAA 120
DB 61 GAACCTTAAGGTCATCCCTGGTTCATAGCAAGGTTTCAGGCACTTCAGCTACATGAA 120
QY 121 CCTTTGTTTGTGTTGTTGTTTAAAGCATTAAATTAATTAATTAAGAGAGTTGG 180
DB 121 CCTTTGTTTGTGTTGTTGTTTAAAGCATTAAATTAATTAATTAAGAGAGTTGG 180
QY 181 CAGTCGTGGCAGACCTTTAATTCAGATATTCAGAGGCGAAGAGGCGAGATCTGT 240
DB 181 CAGTCGTGGCAGACCTTTAATTCAGATATTCAGAGGCGAAGAGGCGAGATCTGT 240
QY 241 GAGTTTCGAAGTCAGCTAGTCTGCAAGGCTAGTCCAGAGTGGCAGAGGCTACACAGAG 300
DB 241 GAGTTTCGAAGTCAGCTAGTCTGCAAGGCTAGTCCAGAGTGGCAGAGGCTACACAGAG 300
QY 301 AACCTTCTCATTAACCAAGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 360
DB 301 AACCTTCTCATTAACCAAGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 360
QY 361 GTCCATTTCAGATGAGCACTCTTAAGATGATCTCTTGACCCAGGTAAAGTAAATGTCA 420
DB 361 GTCCATTTCAGATGAGCACTCTTAAGATGATCTCTTGACCCAGGTAAAGTAAATGTCA 420
QY 421 TGGGGAAGAGGAGTGGAGCTGCTCTAGATTAAAGAGTGAGGCGATGGCTATTCTCA 480
DB 421 TGGGGAAGAGGAGTGGAGCTGCTCTAGATTAAAGAGTGAGGCGATGGCTATTCTCA 480
QY 481 TTTGATTTCATATGAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGAGCTCTGACT 540
DB 481 TTTGATTTCATATGAAAGGCTGATTAAGGCCCAAGAGAGTGAAGTGGAGCTCTGACT 540
QY 541 GAAGACGTGAGCGCTTATTAACACTGGCACTTAAACACTTAAACACTTAAACACTGGCACAG 600
DB 541 GAAGACGTGAGCGCTTATTAACACTGGCACTTAAACACTTAAACACTTAAACACTGGCACAG 600

QY 601 CGTTCAAGGTTTGAAGATCACTTTCAAAACACAGAGAGAAAGTGTGCTGCTCAGC 660
DB 601 CGTTCAAGGTTTGAAGATCACTTTCAAAACACAGAGAGAAAGTGTGCTGCTCAGC 660
QY 661 GTAGCGAGCACTGGCTGCGAGAGAGTGTATTTAGTGAAGAGTACTTCAATATCTTT 720
DB 661 GTAGCGAGCACTGGCTGCGAGAGAGTGTATTTAGTGAAGAGTACTTCAATATCTTT 720
QY 721 GCACCTATCATACAGAGTCAATGTGTACTCCCTAGTCCACAGATGGCTGTAC 780
DB 721 GCACCTATCATACAGAGTCAATGTGTACTCCCTAGTCCACAGATGGCTGTAC 780
QY 781 CTGCTTTCTGCTTTCCCATCTGCTTGCATTTTGCAGAACACAGAAATTAAGTGTGGT 840
DB 781 CTGCTTTCTGCTTTCCCATCTGCTTGCATTTTGCAGAACACAGAAATTAAGTGTGGT 840
QY 841 ATTTATTTGTGTGCTGAGAGACACCATTCAGAGGCTTTTACATTTTCAGGCATGTTTAC 900
DB 841 ATTTATTTGTGTGCTGAGAGACACCATTCAGAGGCTTTTACATTTTCAGGCATGTTTAC 900
QY 901 TAACCTGGCTACTTCTCCCAACGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960
DB 901 TAACCTGGCTACTTCTCCCAACGGTTTGAACCATTTGTTTATATTTACTTATTTGTGT 960
QY 961 GCATGAGGTAGGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 GCATGAGGTAGGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 ATCATTCAGATCCCGACAGCAAGTGAAGTCCAGAGGCTGTGAATTTATGTGGACTG 1080
DB 1021 ATCATTCAGATCCCGACAGCAAGTGAAGTCCAGAGGCTGTGAATTTATGTGGACTG 1080
QY 1081 GGAGCCAAAGGCTGGTCTCTGCAAGAGCAGCAGGCTTAAACATGGAGCACCTCT 1140
DB 1081 GGAGCCAAAGGCTGGTCTCTGCAAGAGCAGCAGGCTTAAACATGGAGCACCTCT 1140
QY 1141 CTAGGCTTAAGTAAATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
DB 1141 CTAGGCTTAAGTAAATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
QY 1201 ACAAGCCTTTAATCCAGCAGCTTGAAGGCTGAGGTGAGGAATTAACACAGGCCAG 1260
DB 1201 ACAAGCCTTTAATCCAGCAGCTTGAAGGCTGAGGTGAGGAATTAACACAGGCCAG 1260
QY 1261 CTGGGGTGCAGAGCTGGCCCTGTTTTTTTGTGTTTCTTATGACACTGGTCTTA 1320
DB 1261 CTGGGGTGCAGAGCTGGCCCTGTTTTTTTGTGTTTCTTATGACACTGGTCTTA 1320
QY 1321 CTTGCTGTATGTCCGTCGAAGGTTGTCAAGTCCCTTGAAGTGAAGTGAAGTGAAGTGAAG 1380
DB 1321 CTTGCTGTATGTCCGTCGAAGGTTGTCAAGTCCCTTGAAGTGAAGTGAAGTGAAGTGAAG 1380
QY 1381 TGATCACGCTGCGTTACAGATCTGGAATTAACCCAGGTGCTCCCTAGAGAGCAGCC 1440
DB 1381 TGATCACGCTGCGTTACAGATCTGGAATTAACCCAGGTGCTCCCTAGAGAGCAGCC 1440
QY 1441 AGTGCCTTTAATCTTGAGCAGCCCTCCCAAGCCTGTTTAAAGACTTTAACTTTTG 1500
DB 1441 AGTGCCTTTAATCTTGAGCAGCCCTCCCAAGCCTGTTTAAAGACTTTAACTTTTG 1500
QY 1501 TGTAAATGTGGAAGTGAAGTGTGCACTTACCAAGTGTGCTGAGCTGAGCATCA 1560
DB 1501 TGTAAATGTGGAAGTGAAGTGTGCACTTACCAAGTGTGCTGAGCATCA 1560
QY 1561 CTGAGCCCGTACCCACAGCACTAGTGAATTAAGGGAACCACTTAACAAATGACA 1620
DB 1561 CTGAGCCCGTACCCACAGCACTAGTGAATTAAGGGAACCACTTAACAAATGACA 1620
QY 1621 ATAGCTTGATAGAGTGTGAATATAGTCTGAGCTATTTGTTACGCTGAGCTTTGCTGTC 1680
DB 1621 ATAGCTTGATAGAGTGTGAATATAGTCTGAGCTATTTGTTACGCTGAGCTTTGCTGTC 1680

Qy 1681 TTAGCATGTGCTGTGAGAGATAGAAAAATGAGACTTGAGTCTGATCTCGGAACCCACA 1740
Db 1681 TTAGCATGTGCTGTGAGAGATAGAAAAATGAGACTTGAGTCTGATCTCGGAACCCACA 1740
Qy 1741 GAGCAGGCGAGAACCCACTCCTGAAAGTTGTTCTGTAGCTTCACATACAACTTCACAT 1800
Db 1741 GAGCAGGCGAGAACCCACTCCTGAAAGTTGTTCTGTAGCTTCACATACAACTTCACAT 1800
Qy 1801 AATAGTTCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Db 1801 AATAGTTCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Qy 1861 GAGATGCGTCAAGCTCCAGAGACACTTGCTGCTCTTGCGAGAGACCTTGATTCAGTTCCC 1920
Db 1861 GAGATGCGTCAAGCTCCAGAGACACTTGCTGCTCTTGCGAGAGACCTTGATTCAGTTCCC 1920
Qy 1921 AGGACTCATATGTTGGCTCAGACGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCCT 1980
Db 1921 AGGACTCATATGTTGGCTCAGACGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCCT 1980
Qy 1981 CTTCTGGGCTCCACAGGACCCACATACATAGTACACACACATACATACATACATACATAC 2040
Db 1981 CTTCTGGGCTCCACAGGACCCACATACATAGTACACACACATACATACATACATACATAC 2040
Qy 2041 CATACACACATTAATTAATTAAGGAACCTTAAAGSTGCATGTGTGTAACATGTGCT 2100
Db 2041 CATACACACATTAATTAATTAAGGAACCTTAAAGSTGCATGTGTGTAACATGTGCT 2100
Qy 2101 TACACATCTGTATGTAAGACATGTACAAACGACACACTGAAGAGGATCTGGGGCTGAG 2160
Db 2101 TACACATCTGTATGTAAGACATGTACAAACGACACACTGAAGAGGATCTGGGGCTGAG 2160
Qy 2161 AGATGGCTCAGGGGTTAAGAGACTGAGACTGCTGCTCCGAGAGAGAGTCTGATCAAT 2220
Db 2161 AGATGGCTCAGGGGTTAAGAGACTGAGACTGCTGCTCCGAGAGAGAGTCTGATCAAT 2220
Qy 2221 CCTAGCAACACATGTTGGCTCACAAACCATCATATAGATCTGACACCCCTCTGCT 2280
Db 2221 CCTAGCAACACATGTTGGCTCACAAACCATCATATAGATCTGACACCCCTCTGCT 2280
Qy 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATCTAATTAATTAATCTTTT 2340
Db 2281 GCATCTGAAGACAGCTGCAGAGCTACAGTGTACTAGATATCTAATTAATTAATCTTTT 2340
Qy 2341 TTTAAAAAATGAAGAGGATCTGAGACACTCAAAAGAGATTAAGACAGTCAAG 2400
Db 2341 TTTAAAAAATGAAGAGGATCTGAGACACTCAAAAGAGATTAAGACAGTCAAG 2400
Qy 2401 GGTGATTAATCTATCCTGGAAGTTTCTTCGCGTTCGCTGCAACTGGTGACAGACG 2460
Db 2401 GGTGATTAATCTATCCTGGAAGTTTCTTCGCGTTCGCTGCAACTGGTGACAGACG 2460
Qy 2461 CCCCCTTTTCATTGACAGAGAGGGGTCTACATTAATTTCTGAACAAAGACACCTGCAGT 2520
Db 2461 CCCCCTTTTCATTGACAGAGAGGGGTCTACATTAATTTCTGAACAAAGACACCTGCAGT 2520
Qy 2521 ATGTTTACTGTCTCTTGCTGACTATAGACAGCGCGCGCGCGCGACACACACACAC 2580
Db 2521 ATGTTTACTGTCTCTTGCTGACTATAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2580
Qy 2581 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAGGTCA 2640
Db 2581 ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAAGGTCA 2640
Qy 2641 AGAAGAGCTGCGCTCAAAACAGATCTTCATCTTCCTCTCAAAAGAGACACAGATTC 2700
Db 2641 AGAAGAGCTGCGCTCAAAACAGATCTTCATCTTCCTCTCAAAAGAGACACAGATTC 2700
Qy 2701 AAGGTGCGAGAAGATCTACAGGGGCGAGAGCGAGGAGGGGAGAGCCATGTTTCC 2760
Db 2701 AAGGTGCGAGAAGATCTACAGGGGCGAGAGCGAGGAGGGGAGAGCCATGTTTCC 2760
Qy 2761 AGAGACTACAGCAGAGGCGAGCAAGATCCCAAGGTCCAGGCGAGGAGGTGAGG 2820

Db 2761 AGAGACTACAGCAGAGGCGAGCAAGCAGATPCCCAAGTCTCAGGCGAGGAGGTGAGG 2820
Qy 2821 CCCCTGTTCCGAGAGAAAGCGAGCGCGAGAAACAGGTTCAAAAGCACAGTTTATGCA 2880
Db 2821 CCCCTGTTCCGAGAGAAAGCGAGCGCGAGAAACAGGTTCAAAAGCACAGTTTATGCA 2880
Qy 2881 GCTCATAAAGTGGAGGTCTGAGGCTCAGTCAAGAAAGAGGAAAGGAAAGGAGGCTTGT 2940
Db 2881 GCTCATAAAGTGGAGGTCTGAGGCTCAGTCAAGAAAGAGGAAAGGAAAGGAGGCTTGT 2940
Qy 2941 GCCCACTGAGAGAGGATCATCTGAGTAGAGAGATCTGCAAGGGGTGCGAGAGGCCCA 3000
Db 2941 GCCCACTGAGAGAGGATCATCTGAGTAGAGAGATCTGCAAGGGGTGCGAGAGGCCCA 3000
Qy 3001 CTGTCTGCCAAGAGAAACCCCAAGTGAACCTGCGCTTGGTGTGATTCACGTA 3060
Db 3001 CTGTCTGCCAAGAGAAACCCCAAGTGAACCTGCGCTTGGTGTGATTCACGTA 3060
Qy 3061 CAAGACCCGAGAGGTCTACATCCCATCCAGTGTGCGCGCGCGCGCACACCCCA 3120
Db 3061 CAAGACCCGAGAGGTCTACATCCCATCCAGTGTGCGCGCGCGCGCGCACACCCCA 3120
Qy 3121 CCCCCGACTCCCGTGCACACTTCTGTAGGGCTGAGAGGTGGCCAGGCCCTGTGGGGTTGC 3180
Db 3121 CCCCCGACTCCCGTGCACACTTCTGTAGGGCTGAGAGGTGGCCAGGCCCTGTGGGGTTGC 3180
Qy 3181 CTACCTGAGTAGAGCCAGGCTCCAGCGGAAAGTGCACCCCATCTCTGAGCTGACGA 3240
Db 3181 CTACCTGAGTAGAGCCAGGCTCCAGCGGAAAGTGCACCCCATCTCTGAGCTGACGA 3240
Qy 3241 GCCAAGGGGGGGGACACAGGAGCTCAGGCTGACAGGCTGTGCGGGCTCTAGATTGCC 3300
Db 3241 GCCAAGGGGGGGGACACAGGAGCTCAGGCTGACAGGCTGTGCGGGCTCTAGATTGCC 3300
Qy 3301 AAGGACCTGGGACCTACTTCCCAACCCGCCCATCATCTCTCTGGGGCCCTATCTTC 3360
Db 3301 AAGGACCTGGGACCTACTTCCCAACCCGCCCATCATCTCTCTGGGGCCCTATCTTC 3360
Qy 3361 CTTATATGTTAAGGAAGTCTCTGGGGGGGGGGTGTGTGTAAGGAGGCTGTTCC 3420
Db 3361 CTTATATGTTAAGGAAGTCTCTGGGGGGGGGGTGTGTGTAAGGAGGCTGTTCC 3420
Qy 3421 GTCCTCTCAGCAGCTTGCACAACTTCTAAGATTCGCCAGGTTGGGCTCTTC 3480
Db 3421 GTCCTCTCAGCAGCTTGCACAACTTCTAAGATTCGCCAGGTTGGGCTCTTC 3480
Qy 3481 CAGACAGCTAAGCAATTTGGTGGGACACATGTGTGACACAGTGTGTGAGGGGACAG 3540
Db 3481 CAGACAGCTAAGCAATTTGGTGGGACACATGTGTGACACAGTGTGTGAGGGGACAG 3540
Qy 3541 GGTCTCTCTCTCTCTGCGACGCTGTCTTCTGTACACCTTGATTAAGTTTGGGG 3600
Db 3541 GGTCTCTCTCTCTCTGCGACGCTGTCTTCTGTACACCTTGATTAAGTTTGGGG 3600
Qy 3601 TGAGGTAAAGTGTCTGAACTCTGAAGAAAGCAAGCAAGCAAGGCTGTGGGGCT 3660
Db 3601 TGAGGTAAAGTGTCTGAACTCTGAAGAAAGCAAGCAAGCAAGGCTGTGGGGCT 3660
Qy 3661 TCAATGAAGAAATTCAAGACACCCCTTCTGTAAGTACACTTCGCTCATCTGTAG 3720
Db 3661 TCAATGAAGAAATTCAAGACACACCCCTTCTGTAAGTACACTTCGCTCATCTGTAG 3720
Qy 3721 ATTCCCTGGACCAAGGTGCTCTGAGACATGATTTCTCAATTAATCAGACAGT 3780
Db 3721 ATTCCCTGGACCAAGGTGCTCTGAGACATGATTTCTCAATTAATCAGACAGT 3780
Qy 3781 CCTGAGATTTGGACTCCGTGCTGATTTACTACTTCTCTGCTGCTCATTTCTGTGT 3840
Db 3781 CCTGAGATTTGGACTCCGTGCTGATTTACTACTTCTCTGCTGCTCATTTCTGTGT 3840
Qy 3841 TCATGCTTTACACATCTGAATGTTTCTTGTGTACACATTCCTCAGACTCTCTGGA 3900
Db 3841 TCATGCTTTACACATCTGAATGTTTCTTGTGTACACATTCCTCAGACTCTCTGGA 3900

Db 3841 TCATGTCCTTACACATCTGAATGATTTCTTTGTGTCACCATTCCTCCCTGACACTCTGGGA 3900
Qy 3901 GATCGTATCCTTGGCAGATGATCTGGGATGTAGCTGAGCCAGCAGAGAGAGGGG 3960
Db 3901 GGTGCTATCCTTGGCAGATGATCTGGGATGTAGCTGAGCCAGCAGAGAGAGGGG 3960
Qy 3961 AGAGTCAGAGCTGTGCTTAGGCCCCATTAGGCCCTGGACATCACCCCTTCTAGAAAT 4020
Db 3961 AGAGTCAGAGCTGTGCTTAGGCCCCATTAGGCCCTGGACATCACCCCTTCTAGAAAT 4020
Qy 4021 GGGCCCCCTATTTTGGGTTACCATGATCTATTTTATATAGAGTGGGCGTAAAGCCA 4080
Db 4021 GGGCCCCCTATTTTGGGTTACCATGATCTATTTTATATAGAGTGGGCGTAAAGCCA 4080
Qy 4081 AACCTGCCAGAAATTTGGGACTCAGCAGACCAAGTTATCTGCTCAGAAATCCCCCTG 4140
Db 4081 AACCTGCCAGAAATTTGGGACTCAGCAGACCAAGTTATCTGCTCAGAAATCCCCCTG 4140
Qy 4141 TCACCTGAGGTTGGGAAATCTGCTCTGGGGGCTTCCAGGCTTGGTTAGCAGAGAGGT 4200
Db 4141 TCACCTGAGGTTGGGAAATCTGCTCTGGGGGCTTCCAGGCTTGGTTAGCAGAGAGGT 4200
Qy 4201 ATCCCTTTGATAGGCGATGACCTAGTCTATGCTGTACTACTATCTCTGCTCAGTTAAAG 4260
Db 4201 ATCCCTTTGATAGGCGATGACCTAGTCTATGCTGTACTACTATCTCTGCTCAGTTAAAG 4260
Qy 4261 CTGGAACTAAACCCAGCGAGCGCCAGGATCTCTACAGTTGTACCCCAAGAACACA 4320
Db 4261 CTGGAACTAAACCCAGCGAGCGCCAGGATCTCTACAGTTGTACCCCAAGAACACA 4320
Qy 4321 AGACAGTATGATGCAAGATAGGTAGCTGGGAGAGAAAGAACTTAAACCCCCCAAG 4380
Db 4321 AGACAGTATGATGCAAGATAGGTAGCTGGGAGAGAAAGAACTTAAACCCCCCAAG 4380
Qy 4381 GCCACAGGTTCCGTTCCCTAGTTCACAAATGCCAGTATGAGTGTCTAGCTATGGGCTG 4440
Db 4381 GCCACAGGTTCCGTTCCCTAGTTCACAAATGCCAGTATGAGTGTCTAGCTATGGGCTG 4440
Qy 4441 TGAGTTGATGCTACAGAGCTAGTGTATGCTGTAGTGTATGATATGATGAGCAC 4500
Db 4441 TGAGTTGATGCTACAGAGCTAGTGTATGCTGTAGTGTATGATATGATGAGCAC 4500
Qy 4501 TTGGGAGGCTGAAGCAGAGAGATTTGCTATATGTTTGAAGCCAGCTGAGCTATAGAGCA 4560
Db 4501 TTGGGAGGCTGAAGCAGAGAGATTTGCTATATGTTTGAAGCCAGCTGAGCTATAGAGCA 4560
Qy 4561 GACTTTGCTTTAAAGAAAAATGAAGAACCCAGAGCTGTGGACACGCTTTAATCCCA 4620
Db 4561 GACTTTGCTTTAAAGAAAAATGAAGAACCCAGAGCTGTGGACACGCTTTAATCCCA 4620
Qy 4621 GCACCTTGGAGGAGAGAGAGAGATTTCTGAGTTCAAGGCCAGCTGGTCTATAGAGT 4680
Db 4621 GCACCTTGGAGGAGAGAGAGAGATTTCTGAGTTCAAGGCCAGCTGGTCTATAGAGT 4680
Qy 4681 GACTTCCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAACCCAGAGAAACAA 4740
Db 4681 GACTTCCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAACCCAGAGAAACAA 4740
Qy 4741 AACAAAAAACAACCAAAACCCAAACCCAAACCCAAACCTCATCTCATCTCTC 4800
Db 4741 AACAAAAAACAACCAAAACCCAAACCCAAACCCAAACCTCATCTCATCTCTC 4800
Qy 4801 TAGGCTGTGCTGTAGTGTAGAGTTTGGGACTTCAAGCTTATATTAATAGGCC 4860
Db 4801 TAGGCTGTGCTGTAGTGTAGAGTTTGGGACTTCAAGCTTATATTAATAGGCC 4860
Qy 4861 TTTTATACACTGTGTCAGAGACGAGAAAGTTTCAAGTCTGGACACAGTGGACCCCTGAGA 4920
Db 4861 TTTTATACACTGTGTCAGAGACGAGAAAGTTTCAAGTCTGGACACAGTGGACCCCTGAGA 4920
Qy 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGAGGCTTCTTGGAGGAAGTGTCCCGAT 4980
Db 4921 AAGTACTCTTGGCAGCCCAAAATTTCTGGAGGCTTCTTGGAGGAAGTGTCCCGAT 4980

Qy 4981 CAGACTACTGTTCTAGAAAGCAGAGAGAGGTTGGAAGAATGTTGGTGACAGACGTT 5040
Db 4981 CAGACTACTGTTCTAGAAAGCAGAGAGAGGTTGGAAGAATGTTGGTGACAGACGTT 5040
Qy 5041 GGAACAGAAAGCAGAGAGGAGGAGCATCAAGATTTCTAGACATGTAGCTACTTTTGGT 5100
Db 5041 GGAACAGAAAGCAGAGAGGAGGAGCATCAAGATTTCTAGACATGTAGCTACTTTTGGT 5100
Qy 5101 TCTCTGGGTGACAGTGTCTCCCGAGGATAGGCTTATAGAAAGGAGCCAGGGGTAGCC 5160
Db 5101 TCTCTGGGTGACAGTGTCTCCCGAGGATAGGCTTATAGAAAGGAGCCAGGGGTAGCC 5160
Qy 5161 AATGATTTCAAGTTGAGGAGACACATCCAGCCAGGCTCTTCTGCAAGCTTAAAGATG 5220
Db 5161 AATGATTTCAAGTTGAGGAGACACATCCAGCCAGGCTCTTCTGCAAGCTTAAAGATG 5220
Qy 5221 AGAGCCCTTAACCTCCCTGAAAGTTTGAAGGAGACAGAGAGCTGAGATCTTCTA 5280
Db 5221 AGAGCCCTTAACCTCCCTGAAAGTTTGAAGGAGACAGAGAGCTGAGATCTTCTA 5280
Qy 5281 GGGTGAAGGAGAGGATCTGCTCTGACCAACATGGCTAGAGCAGAGAGATTTGAGCCAG 5340
Db 5281 GGGTGAAGGAGAGGATCTGCTCTGACCAACATGGCTAGAGCAGAGAGATTTGAGCCAG 5340
Qy 5341 TTACCCCTCAGAACAGCCATCCCTCTTGGCTCTAAGAGAGGCTGGCCCTTTCTGTTT 5400
Db 5341 TTACCCCTCAGAACAGCCATCCCTCTTGGCTCTAAGAGAGGCTGGCCCTTTCTGTTT 5400
Qy 5401 AAGATCTTACTTTTCTTACAGAGAGAGCAGAGCCCTTGTCCCTCCTCTGTTGCTAA 5460
Db 5401 AAGATCTTACTTTTCTTACAGAGAGAGCAGAGCCCTTGTCCCTCCTCTGTTGCTAA 5460
Qy 5461 TTAACACCCCTGTGTACATTTAGTTATTTTACTGTAGTTTGTCTCAGACACTGCA 5520
Db 5461 TTAACACCCCTGTGTACATTTAGTTATTTTACTGTAGTTTGTCTCAGACACTGCA 5520
Qy 5521 TCTGTAGACCTTCTGCTCTTACCTACACAGAGATGAGCCACATTTCTCACCAGAGAG 5580
Db 5521 TCTGTAGACCTTCTGCTCTTACCTACACAGAGATGAGCCACATTTCTCACCAGAGAG 5580
Qy 5581 TGCAGAGAGAGCCTTGAAGAAAGGCTTAACAGTAAACAAAGATGGCCAGATTAACAA 5640
Db 5581 TGCAGAGAGAGCCTTGAAGAAAGGCTTAACAGTAAACAAAGATGGCCAGATTAACAA 5640
Qy 5641 ACTACTATCCTTTGTACCCAAATTTGTTTGTGGAACAGAGAGGCTGTGAGAGTAT 5700
Db 5641 ACTACTATCCTTTGTACCCAAATTTGTTTGTGGAACAGAGAGGCTGTGAGAGTAT 5700
Qy 5701 GT 5760
Db 5701 GT 5760
Qy 5761 CTTGGGGGACTTTTCACTGCTAAAGATATCTGATATTTGGCGCCATGCCAACAGGGTAT 5820
Db 5761 CTTGGGGGACTTTTCACTGCTAAAGATATCTGATATTTGGCGCCATGCCAACAGGGTAT 5820
Qy 5821 TGGGAGAGTCAAGGCTTTCGCAACACAGTAAAGCTCCCAAGTGGATTTGGGCTGAA 5880
Db 5821 TGGGAGAGTCAAGGCTTTCGCAACACAGTAAAGCTCCCAAGTGGATTTGGGCTGAA 5880
Qy 5881 TCACCAAGGGGAGGCTGATCAGAGTGAACATCAAGATTAAGCACCCCTGTGGG 5940
Db 5881 TCACCAAGGGGAGGCTGATCAGAGTGAACATCAAGATTAAGCACCCCTGTGGG 5940
Qy 5941 GCTCAGAGAGGAGTTTCAAGAGGTAAGGCCAAAGCCATTTATTTATCAAGACATGAC 6000
Db 5941 GCTCAGAGAGGAGTTTCAAGAGGTAAGGCCAAAGCCATTTATTTATCAAGACATGAC 6000
Qy 6001 TCAAAATCAAGTCAAGAGAGATTAAGTGAAGATAGGAGGCTGTCAAGTGGGAGCAC 6060
Db 6001 TCAAAATCAAGTCAAGAGAGATTAAGTGAAGATAGGAGGCTGTCAAGTGGGAGCAC 6060

0Y 6061 TGACCTTGACCTTATTAGTCTAGGCGCAAGAGCAGTCAAGAGGGTACTGGTCTTA 6120
|||||
Db 6061 TGACCTTGACCTTATTAGTCTAGGCGCAAGAGCAGTCAAGAGGGTACTGGTCTTA 6120
0Y 6121 CTCAGCTTGAGACAGGCGACGTGGAGAAATGGGTGACCTCCATCCTGATGAGAGAGGCTGAG 6180
|||||
Db 6121 CTCAGCTTGAGACAGGCGACGTGGAGAAATGGGTGACCTCCATCCTGATGAGAGAGGCTGAG 6180
0Y 6181 CACGACCCAGGTCAAGTGTCTCCCTGTGTCTCATGGCAGATTCCTGCGCAAGTTTCAAG 6240
|||||
Db 6181 CACGACCCAGGTCAAGTGTCTCCCTGTGTCTCATGGCAGATTCCTGCGCAAGTTTCAAG 6240
0Y 6241 GACTAAGGACTATCTCTGTGGAAACAAGTATCCAAAGCCCTTAAGCCCATTTTGCTG 6300
|||||
Db 6241 GACTAAGGACTATCTCTGTGGAAACAAGTATCCAAAGCCCTTAAGCCCATTTTGCTG 6300
0Y 6301 AATTAATCAGAACCCCTGGGGATGCGAGGCTCTGAGCAGCAGAGGCTTTTAAAAAGCTC 6360
|||||
Db 6301 AATTAATCAGAACCCCTGGGGATGCGAGGCTCTGAGCAGCAGAGGCTTTTAAAAAGCTC 6360
0Y 6361 CCAGGTGATTTGATCAGACGCTGGAAACAACAGCTCAGGTTTCAACAGAAAGAGGC 6420
|||||
Db 6361 CCAGGTGATTTGATCAGACGCTGGAAACAACAGCTCAGGTTTCAACAGAAAGAGGC 6420
0Y 6421 AAAGCTAAGGAAAGCTTGGGATGGGGAGGCTTCTTCAGGCGCAGTATGAGGGCTGGT 6480
|||||
Db 6421 AAAGCTAAGGAAAGCTTGGGATGGGGAGGCTTCTTCAGGCGCAGTATGAGGGCTGGT 6480
0Y 6481 AGCAGTGTGGAGGCTTCTCTGCGCTGTCAATATAGCTATCAGTCACTCATCCATCCAT 6540
|||||
Db 6481 AGCAGTGTGGAGGCTTCTCTGCGCTGTCAATATAGCTATCAGTCACTCATCCATCCAT 6540
0Y 6541 ACACCCAGCCATTCATTTATGCAACCCATCCCTTCATCCATCCATCTATCCAGTACCCAC 6600
|||||
Db 6541 ACACCCAGCCATTCATTTATGCAACCCATCCCTTCATCCATCCATCTATCCAGTACCCAC 6600
0Y 6601 CCAGGATCATCCAAACCT 6660
|||||
Db 6601 CCAGGATCATCCAAACCT 6660
0Y 6661 CATTTATCCAAAGAGAACTGTGTATGTACTAAATGTGGAGAAATTTAATTTTATAGA 6720
|||||
Db 6661 CATTTATCCAAAGAGAACTGTGTATGTACTAAATGTGGAGAAATTTAATTTTATAGA 6720
0Y 6721 ACCTCTGTTGATGACTGATGTTGTGATGTATGTGACAGGTACATACCAAGCACAGCG 6780
|||||
Db 6721 ACCTCTGTTGATGACTGATGTTGTGATGTATGTGACAGGTACATACCAAGCACAGCG 6780
0Y 6781 TGGCAATCGSAGAAAGGTTTGGGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
|||||
Db 6781 TGGCAATCGSAGAAAGGTTTGGGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
0Y 6841 TTGAACCTAAATTTATCGGGCTGTGGCAAGTGTCTTTACCAAGCAGACCAATTTTGTGACA 6900
|||||
Db 6841 TTGAACCTAAATTTATCGGGCTGTGGCAAGTGTCTTTACCAAGCAGACCAATTTTGTGACA 6900
0Y 6901 CATCATTTATTATTAGAAAGCATCTTATGTATGTGACAGGCTGAGCTCAAGCTTGCTATGTG 6960
|||||
Db 6901 CATCATTTATTATTAGAAAGCATCTTATGTATGTGACAGGCTGAGCTCAAGCTTGCTATGTG 6960
0Y 6961 CCACGGATGACCTTTTAACCTGCTGCTCTTCAGGCTCCACCGAGCTCTAGGTTTATAGGT 7020
|||||
Db 6961 CCACGGATGACCTTTTAACCTGCTGCTCTTCAGGCTCCACCGAGCTCTAGGTTTATAGGT 7020
0Y 7021 GTTCAACTGTGTAATGCTTTAATCCACAGCTGTGTGGGGGGGGGGGGAGGCGGATTC 7080
|||||
Db 7021 GTTCAACTGTGTAATGCTTTAATCCACAGCTGTGTGGGGGGGGGGGGAGGCGGATTC 7080
0Y 7081 CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGATACCTTGGGGCTATACAGGGAA 7140
|||||
Db 7081 CCTGAGTTGGAGGCCAGTTTGGTCTACAGAGTTTCAGATACCTTGGGGCTATACAGGGAA 7140
0Y 7141 ACCCTATCCCAAAACAACAACAACAACAACAATAATTTCTGTGCAATATACAGAGCA 7200

Db 7141 ACCCTATCCCAAAACAACAACAACAACAACAATAATTTCTGTGCAATATACAGAGCA 7200
|||||
0Y 7201 TTAGAGGATATTAGTAGGCTAGGCGTGTGTAGGAGAGTCACTCTTTCTTTATTT 7260
|||||
Db 7201 TTAGAGGATATTAGTAGGCTAGGCGTGTGTAGGAGAGTCACTCTTTCTTTATTT 7260
0Y 7261 ATTAATGTAAAGTACTCAACAAGTATGATATATCTATCTATCTATCTATCTATCTATC 7320
|||||
Db 7261 ATTAATGTAAAGTACTCAACAAGTATGATATATCTATCTATCTATCTATCTATCTATC 7320
0Y 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380
|||||
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380
0Y 7381 CTGCTTTGACTGTGAATGCTCTTATTTCTGGGGTCAACTCTCAACCCCTAGTGGGTTT 7440
|||||
Db 7381 CTGCTTTGACTGTGAATGCTCTTATTTCTGGGGTCAACTCTCAACCCCTAGTGGGTTT 7440
0Y 7441 ACCAACAACCAGACATTTATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTT 7500
|||||
Db 7441 ACCAACAACCAGACATTTATTTATTTGTTTATTTATTTATTTATTTATTTATTTATTT 7500
0Y 7501 GGACTGAGGCTCTTGTGATGCTTAAGCAAGCTCTGCGACAGAGCTGCGAGCTCCAGTCC 7560
|||||
Db 7501 GGACTGAGGCTCTTGTGATGCTTAAGCAAGCTCTGCGACAGAGCTGCGAGCTCCAGTCC 7560
0Y 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTGTCAATATTCGACAGCGCTATGTAGCTCTC 7620
|||||
Db 7561 CCATTTTGTTCAGGTGACTCTGTGACAGTGTGTCAATATTCGACAGCGCTATGTAGCTCTC 7620
0Y 7621 CACCTCCAGTTCCAGACTTTTCTGGTCAATCCAGTGGGGCGGGCAACTCTGTGCTTACCA 7680
|||||
Db 7621 CACCTCCAGTTCCAGACTTTTCTGGTCAATCCAGTGGGGCGGGCAACTCTGTGCTTACCA 7680
0Y 7681 GTGCGCTGTCCCTGTCTCTCAGACCTACATATTTGGCTCTGTAAGAGTTTCATGTAAATG 7740
|||||
Db 7681 GTGCGCTGTCCCTGTCTCTCAGACCTACATATTTGGCTCTGTAAGAGTTTCATGTAAATG 7740
0Y 7741 GGAATGCGTCTGTGTATTTCTTTATGAGTGGCGCCCTTTATCTTTAGCACAGTTTGTGTTG 7800
|||||
Db 7741 GGAATGCGTCTGTGTATTTCTTTATGAGTGGCGCCCTTTATCTTTAGCACAGTTTGTGTTG 7800
0Y 7801 GGGCATGTGTCACTGTATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7860
|||||
Db 7801 GGGCATGTGTCACTGTATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7860
0Y 7861 TGTGTGATTAACACATTTCTGTTCATTTACTGATGGAATTTGTGGCCCAACCCGCAC 7920
|||||
Db 7861 TGTGTGATTAACACATTTCTGTTCATTTACTGATGGAATTTGTGGCCCAACCCGCAC 7920
0Y 7921 CCTTTTCTTTTATTTTGAAGCAAGTCTTGTGTAACTTTCGAATCTGGCTGTCT 7980
|||||
Db 7921 CCTTTTCTTTTATTTTGAAGCAAGTCTTGTGTAACTTTCGAATCTGGCTGTCT 7980
0Y 7981 GAGCTACCTGTGTAGACAGGCTGTAGGCTGTCTTCCACTTTTACACTCTCTGTGAAC 8040
|||||
Db 7981 GAGCTACCTGTGTAGACAGGCTGTAGGCTGTCTTCCACTTTTACACTCTCTGTGAAC 8040
0Y 8041 AGAGTACCATGACATCTTCAAGACATTTTCTGTGTGTTTGTGTTTATACATTTGTG 8100
|||||
Db 8041 AGAGTACCATGACATCTTCAAGACATTTTCTGTGTGTTTGTGTTTATACATTTGTG 8100
0Y 8101 TGTATGCTGTATATGTGATGTTGTGTCTTCAGGCTGCAATGTGTGCTGTG 8160
|||||
Db 8101 TGTATGCTGTATATGTGATGTTGTGTCTTCAGGCTGCAATGTGTGCTGTG 8160
0Y 8161 TGGGACAGAGAACCAACCATGTGCCATTCCTAGATACGCACTGTGTATATATGTA 8220
|||||
Db 8161 TGGGACAGAGAACCAACCATGTGCCATTCCTAGATACGCACTGTGTATATATGTA 8220
0Y 8221 TGTATATGTATGTATTTATTTAGTGTGCCAAGTATCAGGTATTTTGTGAGTTTTCAC 8280
|||||

| | | | |
|----------|-------------|--|------|
| Db | 8221 | TGTTATTTAGATGTTTATTTATTAAGTGTGCCCAAGATATGACAGTATTTTGTGGAGTTTTCAC | 8280 |
| Oy | 8281 | CTTCCCTTTGGGCTCTCCGCAATTAACCTAGCTCTCGGGCTAGTGAACAATGCTTCCA | 8340 |
| Db | 8281 | CTTCCCTTTGGGCTCTCCGCAATTAACCTAGCTCTCGGGCTAGTGAACAATGCTTCCA | 8340 |
| Oy | 8341 | CTGCATGAGCAATCTGCCTGCTCCCTGTGCACCTCTCTTATTTTCCAGATGGGACTA | 8400 |
| Db | 8341 | CTGCATGAGCAATCTGCCTGCTCCCTGTGCACCTCTCTTATTTTCCAGATGGGACTA | 8400 |
| Oy | 8401 | CGCACTGCACCTGGCTTAAAGCTCACCAGTCAATCCAGATGGCTTAGCGAGAGACTAG | 8460 |
| Db | 8401 | CGCACTGCACCTGGCTTAAAGCTCACCAGTCAATCCAGATGGCTTAGCGAGAGACTAG | 8460 |
| Oy | 8461 | GGATATGCTGGCTCTCTCCCTCCACAGTGTAGATTTACAGGCATATACCTACTGCTGGAG | 8520 |
| Db | 8461 | GGATATGCTGGCTCTCTCCCTCCACAGTGTAGATTTACAGGCATATACCTACTGCTGGAG | 8520 |
| Oy | 8521 | ATTTTAAACCGAATCTGTGAGATAGGAGGACTCTACCAATGAGAGGCTCTTTTGT | 8580 |
| Db | 8521 | ATTTTAAACCGAATCTGTGAGATAGGAGGACTCTACCAATGAGAGGCTCTTTTGT | 8580 |
| Oy | 8581 | GTTTGGTTTGGTTTCTCTGCTCATAGATCAGGCACTGTGAATTAAGTCTGAGCCTGGCTAC | 8640 |
| Db | 8581 | GTTTGGTTTGGTTTCTCTGCTCATAGATCAGGCACTGTGAATTAAGTCTGAGCCTGGCTAC | 8640 |
| Oy | 8641 | ATTAACATCTTGTCTCAAAAGCCTATAGAGGTAGGAGGTGAGGCTTAAGAGAGCCTT | 8700 |
| Db | 8641 | ATTAACATCTTGTCTCAAAAGCCTATAGAGGTAGGAGGTGAGGCTTAAGAGAGCCTT | 8700 |
| Oy | 8701 | AAGCGGCTGTGATAGACACAGAGATAGGCTGACTATATTAACAAGACCTGTGTTCAAA | 8760 |
| Db | 8701 | AAGCGGCTGTGATAGACACAGAGATAGGCTGACTATATTAACAAGACCTGTGTTCAAA | 8760 |
| Oy | 8761 | ACATGAGAGGAGGGGTATGTTTTAAGTGTGCTGGCTGTGAACAGGCACCTAAGAGGCCAA | 8820 |
| Db | 8761 | ACATGAGAGGAGGGGTATGTTTTAAGTGTGCTGGCTGTGAACAGGCACCTAAGAGGCCAA | 8820 |
| Oy | 8821 | TGTAGACATTTGACTTAAGAAAGATCATCATCAAGCCGGTGGCGAGGTTAGAGTTGG | 8880 |
| Db | 8821 | TGTAGACATTTGACTTAAGAAAGATCATCATCAAGCCGGTGGCGAGGTTAGAGTTGG | 8880 |
| Oy | 8881 | ACTACAGTGGTCAAGACCCCATAGAGAGGAGTTCCTTCTCTGCGGGCTCTCAAGC | 8940 |
| Db | 8881 | ACTACAGTGGTCAAGACCCCATAGAGAGGAGTTCCTTCTCTGCGGGCTCTCAAGC | 8940 |
| Oy | 8941 | CTGGCTGCAGGCGCACTGCTCTCATATGCTTCTCCTTAGGCTGTCACCATG | 8995 |
| Db | 8941 | CTGGCTGCAGGCGCACTGCTCTCATATGCTTCTCCTTAGGCTGTCACCATG | 8995 |
| RESULT 2 | | | |
| AAA49605 | ID | AAA49605 standard; DNA: 8993 bp. | |
| NC | AAA49605: | | |
| XX | 19-DEC-2000 | (first entry) | |
| DT | | | |
| XX | | | |
| DE | | Mouse villin gene regulatory elements and partial sequence. | |
| XX | | | |
| KM | | Mouse; villin; intestinal epithelial cell; | |
| KW | | uro-genital tract epithelial cell; tumour; ds. | |
| XX | | | |
| OS | | Mus sp. | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | misc_signal | 3442 | |
| FT | | /*tag= a | |
| FT | exon | 3442..3487 | |
| FT | | /*tag= b | |
| FT | | /number= 1 | |

| | | |
|----|--|---|
| FT | intron | 3488..8981 |
| FT | | /tag= c |
| FT | exon | /number=1 |
| FT | | 8982..8995 |
| FT | | /tag= d |
| FT | | /number= 2 |
| FT | misc_signal | 8992 |
| FT | | /tag= e |
| FT | | /note= "translational start site" |
| XX | | |
| PN | MOZ00034493-A2. | |
| PD | 15-JUN-2000. | |
| XX | | |
| PF | 09-DEC-1999; | 99WO-EP09782. |
| XX | | |
| PR | 09-DEC-1998; | 98WO-EP08009. |
| XX | | |
| PA | (CNRS) CENT NAT RECH SCI. | |
| XX | (CURT-) INST CURIE. | |
| PI | Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J; | |
| DR | WPI; 2000-423434/36. | |
| XX | | |
| PT | Novel nucleotide sequence derived from mouse villin gene for targeted | |
| PT | expression of transgenes in immature and differentiated epithelial | |
| PT | cells of intestine or urogenital tracts - | |
| XX | | |
| PS | Claim 3; Fig 6; 52pp; English. | |
| CC | The present sequence consists of the regulatory region and the first | |
| CC | exon and intron of the murine villin gene. This gene is expressed in the | |
| CC | epithelial cells of the intestine and uro-genital tracts. Its promoter | |
| CC | sequence can be used in the targeted expression of exogenous genes in | |
| CC | these places, which may, for example, be useful in the treatment of | |
| CC | tumours. | |
| XX | | |
| XQ | Sequence 8993 BP; 2274 A; 2105 C; 2257 G; 2355 T; 2 other: | |
| SO | | |
| | Query Match | 99.7%; Score 8969.6; DB 21; Length 8993; |
| | Best Local Similarity | 100.0%; Pred. No. 0; |
| | Matches 8991; Conservative | 1; Mismatches 1; Indels 2; Gaps 2. |
| OY | 1 | GATCTGGTCACCAAGACACTGTGTCACCACACTGGGAGGTGAGGAGAGGTCA 60 |
| Db | 1 | GATCTGGTCACCAAGACACTGTGTCACCACACTGGGAGGTGAGGAGAGGTCA 60 |
| OY | 61 | GAACTTTAAGGCATCCTTGTTACATAGCAAGGTTTCAGCCAGCTTCAGCTAACNTGAAA 120 |
| Db | 61 | GAACTTTAAGGCATCCTTGTTACATAGCAAGGTTTCAGCCAGCTTCAGCTAACNTGAAA 120 |
| OY | 121 | CCTTGTGTTGTTGTTGTTGTTGTTTAAGAAGCATTAATAATACCAATGAAGAGGTTGG 180 |
| Db | 121 | CCTTGTGTTGTTGTTGTTGTTGTTTAAGAAGCATTAATAATACCAATGAAGAGGTTGG 180 |
| OY | 181 | CAGTGTGGCAGACACCTTTAATTCAGTAGTATTCAGSAGGACAGACAGGACAGATCTGT 240 |
| Db | 181 | CAGTGTGGCAGACACCTTTAATTCAGTAGTATTCAGSAGGACAGACAGGACAGATCTGT 240 |
| OY | 241 | GAGTTCGAAAGTCAGGCTAGTCGCAAACTTAATTCAGAGTGGCAAAGGCTTCACAGAGA 300 |
| Db | 241 | GAGTTCGAAAGTCAGGCTAGTCGCAAACTTAATTCAGAGTGGCAAAGGCTTCACAGAGA 300 |
| OY | 301 | AACCTTGTCATAAACAAGTAGTAGTAGTAGTAATGCCATAGAGAAAAATTGGA 360 |
| Db | 301 | AACCTTGTCATAAACAAGTAGTAGTAGTAGTAATGCCATAGAGAAAAATTGGA 360 |
| OY | 361 | GTCATTAGAGTAGGACCATCTATTAAGATGATTCCTTGACCCAGGTAAGCTAAATGTCA 420 |
| Db | 361 | GTCATTAGAGTAGGACCATCTATTAAGATGATTCCTTGACCCAGGTAAGCTAAATGTCA 420 |
| OY | 421 | TGGGGAAGGGGAGTGGACTGCTAGATTAAAAAGTCTGAGCGGAGTCTTAATTCCTCAA 480 |

|||||
Db 421 TGGGAAAGGGAGTGGAGCTGTCTAGATTAAAGTGTGAGGAGTCTATTCTCAA 480
Oy 481 TTTGATTTCCATATGAAAAGCGTGAATAGGCCCAAGAGAAGTGGAACTGTGGACT 540
Db 481 TTTGATTTCCATATGAAAAGCGTGAATAGGCCCAAGAGAAGTGGAACTGTGGACT 540
Oy 541 GAAGACGTGACGGCCCTTTAAACACTGGACCTTATAACACTTATAAACACTGGACAGG 600
Db 541 GAAGACGTGACGGCCCTTTATAACACTGGACCTTATAACACTTATAAACACTGGACAGG 600
Oy 601 CGTTGAGGTTTGAAAGATCACTTTCAAAACACAGAAAGAGTGGTCCGCTCCCTCAGC 660
Db 601 CGTTGAGGTTTGAAAGATCACTTTCAAAACACAGAAAGAGTGGTCCGCTCCCTCAGC 660
Oy 661 GTAGCGAGCAGTGGCTGGAGAGAGTGAATATTAGTGAAGCTACCTTCACAATATCTTT 720
Db 661 GTAGCGAGCAGTGGCTGGAGAGAGTGAATATTAGTGAAGCTACCTTCACAATATCTTT 720
Oy 721 GCACCTATCACATACAGGTGCAAAATGTGCTACCTCCAGTCCACAGATGGCTGTACA 780
Db 721 GCACCTATCACATACAGGTGCAAAATGTGCTACCTCCAGTCCACAGATGGCTGTACA 780
Oy 781 CTCGTTTCTGCTTTCCCATCTGTGTGACATTTGTGAGAACCCAGAATTTAGAATGTGGT 840
Db 781 CTCGTTTCTGCTTTCCCATCTGTGTGACATTTGTGAGAACCCAGAATTTAGAATGTGGT 840
Oy 841 ATTTATTTGTGTGTGAGAGACACCATCCAGGGCTTTTCACATTCAGCAGCATGTGTTAC 900
Db 841 ATTTATTTGTGTGTGAGAGACACCATCCAGGGCTTTTCACATTCAGCAGCATGTGTTAC 900
Oy 901 TAACGGGGCTACTTCCACAGGCTTGAACCATTTGTTTTATTTACTTATTTGTGT 960
Db 901 TAACGGGGCTACTTCCACAGGCTTGAACCATTTGTTTTATTTACTTATTTGTGT 960
Oy 961 GCATGAGGTAGGACATGATACGTATAGTATAGAGTCAATGATGTGGCTGCTACCTCAA 1020
Db 961 GCATGAGGTAGGACATGATACGTATAGTATAGAGTCAATGATGTGGCTGCTACCTCAA 1020
Oy 1021 ATCATTTGACAGATCCCACAGCAAGTGAAGTCCAGACGCTTTGTAGTTGTTATGTGGACTG 1080
Db 1021 ATCATTTGACAGATCCCACAGCAAGTGAAGTCCAGACGCTTTGTAGTTGTTATGTGGACTG 1080
Oy 1081 GGAGCGCAAGGCTGGGTTGCTGCAAGAGAGCCAGTGGGCTTAACATGGGACAGCTCT 1140
Db 1081 GGAGCGCAAGGCTGGGTTGCTGCAAGAGAGCCAGTGGGCTTAACATGGGACAGCTCT 1140
Oy 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAATAATATATTCTCAGCGGCTGTGGTGC 1200
Db 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAATAATATATTCTCAGCGGCTGTGGTGC 1200
Oy 1201 ACAGCGCTTTAATCCCACAGCTTGAGAGGCTGAGGTGTAGCAATTAACACAGGCCAG 1260
Db 1201 ACAGCGCTTTAATCCCACAGCTTGAGAGGCTGAGGTGTAGCAATTAACACAGGCCAG 1260
Oy 1261 CTGGGGTGCAGAGCTTTGGCCCTGTTTTTTTGTCTTTTCTTATGTGACCTGGTGTCTTA 1320
Db 1261 CTGGGGTGCAGAGCTTTGGCCCTGTTTTTTTGTCTTTTCTTATGTGACCTGGTGTCTTA 1320
Oy 1321 CCTGCGTATATGTCGGTCAAGGGTGTAGATCCCTTGGAGCTGAGTTAAAGACAGTTG 1380
Db 1321 CCTGCGTATATGTCGGTCAAGGGTGTAGATCCCTTGGAGCTGAGTTAAAGACAGTTG 1380
Oy 1381 TGATACAGCTGTGCGTTACAGATGCTGGAATTTGAACCCAGGTGTCCCTAGAGACAGCC 1440
Db 1381 TGATACAGCTGTGCGTTACAGATGCTGGAATTTGAACCCAGGTGTCCCTAGAGACAGCC 1440
Oy 1441 AGTGTCTTAACTTCTGAGCCACCCCTCCAAACCTTGTCTTTTGAAGACTTTAACTTTTG 1500
Db 1441 AGTGTCTTAACTTCTGAGCCACCCCTCCAAACCTTGTCTTTTGAAGACTTTAACTTTTG 1500
Oy 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTAACAAAGTGTGTGCTGGCTGTAGCATCA 1560
Db 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTAACAAAGTGTGTGCTGGCTGTAGCATCA 1560

Db 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTAACAAAGTGTGTGCTGGCTGTAGCATCA 1560
Oy 1561 CTGAGCCCGCTACCACACAGCACTAGTGTGATACAGTTTAAGGGCAACACTTAACATGACA 1620
Db 1561 CTGAGCCCGCTACCACACAGCACTAGTGTGATACAGTTTAAGGGCAACACTTAACATGACA 1620
Oy 1621 ATAGTTGGATAGATTTGAAATATAGCTCCGACCTATTTGGTTAGCCGTGGACCTTGGCTGCC 1680
Db 1621 ATAGTTGGATAGATTTGAAATATAGCTCCGACCTATTTGGTTAGCCGTGGACCTTGGCTGCC 1680
Oy 1681 TTAGCATGTGCTGTGAGAGATAGAAAAATGAAGACTTGTAGTCTAGCTGGAAACCCACA 1740
Db 1681 TTAGCATGTGCTGTGAGAGATAGAAAAATGAAGACTTGTAGTCTAGCTGGAAACCCACA 1740
Oy 1741 GAGGCGAGCGAGAACCCACTCCTGAAAGTGTGTTCTGTGAGCTTCACATTAACATTCACAT 1800
Db 1741 GAGGCGAGCGAGAACCCACTCCTGAAAGTGTGTTCTGTGAGCTTCACATTAACATTCACAT 1800
Oy 1801 AATAGTTACAAATGATTAATTAATTAATTAATTTCTTTTAAAGGTATATGTTGGGAGGA 1860
Db 1801 AATAGTTACAAATGATTAATTAATTAATTAATTTCTTTTAAAGGTATATGTTGGGAGGA 1860
Oy 1861 GAGATGGCTCACTCCAGAGACACTTGTGCTCTTGGCAGAGACCTAGACTTCACTTCCC 1920
Db 1861 GAGATGGCTCACTCCAGAGACACTTGTGCTCTTGGCAGAGACCTAGACTTCACTTCCC 1920
Oy 1921 AGGACTCATATGTGTGGCTCACAGCCATCTGTAAATCCATTCACAGGGTTCCACACCT 1980
Db 1921 AGGACTCATATGTGTGGCTCACAGCCATCTGTAAATCCATTCACAGGGTTCCACACCT 1980
Oy 1981 CTTTGGGCTCCACAGGCGCCACATATAGTATAGTACAGACATATATGCGGCAAAAACAC 2040
Db 1981 CTTTGGGCTCCACAGGCGCCACATATAGTATAGTACAGACATATATGCGGCAAAAACAC 2040
Oy 2041 CATACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100
Db 2041 CATACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2100
Oy 2101 TACACATGCTGATTTGAAGACATGTACAAACGACACACATGAAAGGATCTGGGGCTGGAG 2160
Db 2101 TACACATGCTGATTTGAAGACATGTACAAACGACACACATGAAAGGATCTGGGGCTGGAG 2160
Oy 2161 AGATGGCTCAGGCGTTTAAGAGCAGTGCCTTCCTCCGAAGGAAGTCTGTGATTCAAAT 2220
Db 2161 AGATGGCTCAGGCGTTTAAGAGCAGTGCCTTCCTCCGAAGGAAGTCTGTGATTCAAAT 2220
Oy 2221 CCTAGCAACACATGTTGGCTCACAAACATCCATTAATGATCTGACACCCCTCTTGTGT 2280
Db 2221 CCTAGCAACACATGTTGGCTCACAAACATCCATTAATGATCTGACACCCCTCTTGTGT 2280
Oy 2281 GCATCTGAAGACAGCTGCAGAGCTTACAGTGTACTTAAGATTAATTAATTAATTAATTAAT 2340
Db 2281 GCATCTGAAGACAGCTGCAGAGCTTACAGTGTACTTAAGATTAATTAATTAATTAATTAAT 2340
Oy 2341 TTTAAAAAATGAAGGAGATCTGAGACACCTCAAAAGAGATTATAGAGCAGTGCACAG 2400
Db 2341 TTTAAAAAATGAAGGAGATCTGAGACACCTCAAAAGAGATTATAGAGCAGTGCACAG 2400
Oy 2401 GGTGATTAATCTATGCTGAGATTTTCTTTCCGCTTGGCTTGCACACTGGGTGAGACAGG 2460
Db 2401 GGTGATTAATCTATGCTGAGATTTTCTTTCCGCTTGGCTTGCACACTGGGTGAGACAGG 2460
Oy 2461 CCCCCTTTTCAATTCACAAAGAGCGGTGTACATTAATTTTGAACAAAGCAGCCTGCAT 2520
Db 2461 CCCCCTTTTCAATTCACAAAGAGCGGTGTACATTAATTTTGAACAAAGCAGCCTGCAT 2520
Oy 2521 ATGTTTACGCTCTGCTGACTATAGACAGCGGACGCGCGCGGCGCACACACACACACAC 2580
Db 2521 ATGTTTACGCTCTGCTGACTATAGACAGCGGACGCGCGCGGCGCACACACACACACAC 2580
Oy 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA 2640
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGGTCA 2640

| | | | |
|----|-------|---|------|
| QY | 264.1 | AGAGAGGCTGCCTCAACACGATCTCATCTTTCCTCTCTAAAGSAGACCAGGATTC | 2700 |
| Db | 264.1 | AGAGAGGCTGCCTCAACACGATCTCATCTTTCCTCTCTAAAGSAGACCAGGATTC | 2700 |
| QY | 270.1 | AGAGGCGCAGAAAGATCTCAAGAGGGGGGCGAGAGCCGAGAGGGGGGAGACAGGCCATGGTTTCC | 2760 |
| Db | 270.1 | AGAGGCGCAGAAAGATCTCAAGAGGGGGGCGAGAGCCGAGAGGGGGGAGACAGGCCATGGTTTCC | 2760 |
| QY | 276.1 | AGAGACCTACAGCAGAGAGGGGAGCAAGCAGATCCCCAGGTCCAGGGCAGGGAGGTGAGG | 2820 |
| Db | 276.1 | AGAGACCTACAGCAGAGAGGGGAGCAAGCAGATCCCCAGGTCCAGGGCAGGGAGGTGAGG | 2820 |
| QY | 282.1 | CCCTTTGTTCCGAGAGAGAGCCAGCGGACAGACAGGTTTCAAGGCACAGTTTATGSCA | 2880 |
| Db | 282.1 | CCCTTTGTTCCGAGAGAGAGCCAGCGGACAGACAGGTTTCAAGGCACAGTTTATGSCA | 2880 |
| QY | 288.1 | GCTCATTAAGCTGGAGGTCCGTGGCTCATCTCAGAAAGAGGAAGAGGAAGGCCCTTGT | 2940 |
| Db | 288.1 | GCTCATTAAGCTGGAGGTCCGTGGCTCATCTCAGAAAGAGGAAGAGGAAGGCCCTTGT | 2940 |
| QY | 294.1 | GCCCACTAGAGGAGGTGATGCTGAGTAGAGAGATCTCAGGGGTGCAGAGAGCCCA | 3000 |
| Db | 294.1 | GCCCACTAGAGGAGGTGATGCTGAGTAGAGAGATCTCAGGGGTGCAGAGAGCCCA | 3000 |
| QY | 300.1 | CTGTCTGTCCCAAGGAACCCCAAGTGTGAATCTTGCCCTTGAGTTCAGCTA | 3060 |
| Db | 300.1 | CTGTCTGTCCCAAGGAACCCCAAGTGTGAATCTTGCCCTTGAGTTCAGCTA | 3060 |
| QY | 306.1 | CAAGACCCCAAGAGTCTTACTCATCCCATTCAGTGCCTCCGCCGCCACACCCCA | 3120 |
| Db | 306.1 | CAAGACCCCAAGAGTCTTACTCATCCCATTCAGTGCCTCCGCCGCCACACCCCA | 3120 |
| QY | 312.1 | CCCCGACTCCCGGCCACTTCTCTAGGGGTGAGGGTGGCCAGCCCTGAGGGGGTTGC | 3180 |
| Db | 312.1 | CCCCGACTCCCGGCCACTTCTCTAGGGGTGAGGGTGGCCAGCCCTGAGGGGGTTGC | 3180 |
| QY | 318.1 | CTACCTGAGGTAGAGCCAGGTCCTAGGCCGGAAGTGCACCCCATCTCTGAAGCTGCAGA | 3240 |
| Db | 318.1 | CTACCTGAGGTAGAGCCAGGTCCTAGGCCGGAAGTGCACCCCATCTCTGAAGCTGCAGA | 3240 |
| QY | 324.1 | GCCAAAGGGGGGCGACACGGCAGCTCAGGCTGTGAGGCTGTAGGCTTCC | 3300 |
| Db | 324.1 | GCCAAAGGGGGGCGACACGGCAGCTCAGGCTGTGAGGCTGTAGGCTTCC | 3300 |
| QY | 330.1 | AGGGACCGGGGACACTCTTCCCGCACCCGCCATTCATCTCTGAGGGGCCCTATCTTCC | 3360 |
| Db | 330.1 | AGGGACCGGGGACACTCTTCCCGCACCCGCCATTCATCTCTGAGGGGCCCTATCTTCC | 3360 |
| QY | 336.1 | CTTATTATGTGAAGGAAGTTCCTGAGGGGGGGGGGCTGTGTGAGAGCAAAAGTGTTCG | 3420 |
| Db | 336.1 | CTTATTATGTGAAGGAAGTTCCTGAGGGGGGGGGGCTGTGTGAGAGCAAAAGTGTTCG | 3420 |
| QY | 342.1 | GTTCTCTGACCCAGCTTGGCCACACTTCTTAAGATCTCCAGAGGTGGCTCTTTC | 3480 |
| Db | 342.1 | GTTCTCTGACCCAGCTTGGCCACACTTCTTAAGATCTCCAGAGGTGGCTCTTTC | 3480 |
| QY | 348.1 | CAGACAGGTAAAGCAATTTGGGTGGGACACATGTGTACACAGGTGTGGAGGGGACAG | 3540 |
| Db | 348.1 | CAGACAGGTAAAGCAATTTGGGTGGGACACATGTGTACACAGGTGTGGAGGGGACAG | 3540 |
| QY | 354.1 | GGTCTTGTCTCTCTGAGAGCTGTGCTTCTGTAGCACTTGTATAGTTTGGGGG | 3600 |
| Db | 354.1 | GGTCTTGTCTCTCTGAGAGCTGTGCTTCTGTAGCACTTGTATAGTTTGGGGG | 3600 |
| QY | 360.1 | TGAGGTAAAGTGTCTTGAACCTCTGAAGAAGCAAGAACCCAGCAGGCTCTTGGGCT | 3660 |
| Db | 360.1 | TGAGGTAAAGTGTCTTGAACCTCTGAAGAAGCAAGAACCCAGCAGGCTCTTGGGCT | 3660 |
| QY | 366.1 | TCATATGAAGGAAGTTCACACACCCCTTCTGTAAAGTACACTTGGCTTCACTGTGTAG | 3720 |
| Db | 366.1 | TCATATGAAGGAAGTTCACACACCCCTTCTGTAAAGTACACTTGGCTTCACTGTGTAG | 3720 |

| | | | |
|----|------|--|------|
| OY | 3721 | ATTCCTGGGACCAAGTGGCTCCTGGGACATGAGATTTCTACAATTTAAATTCAGGACAGT | 3780 |
| Dp | 3721 | ATTCCTGGGACCAAGTGGCTCCTGGGACATGAGATTTCTACAATTTAAATTCAGGACAGT | 3780 |
| OY | 3781 | CCTGAGACTTGGACCTCCGCTCGTATTTATACATCTCTCGGCTGCATATTTCTGTGT | 3840 |
| Dp | 3781 | CCTGAGACTTGGACCTCCGCTCGTATTTATACATCTCTCGGCTGCATATTTCTGTGT | 3840 |
| OY | 3841 | TCATGTCTTACACATCTGAAATGGTTTCTTGTGTGTCACATTTCCCTGACACTCCTGGGA | 3900 |
| Dp | 3841 | TCATGTCTTACACATCTGAAATGGTTTCTTGTGTGTCACATTTCCCTGACACTCCTGGGA | 3900 |
| OY | 3901 | GGTGTATCTTGGCACATATATCTCTGGGATGTAAGCTCACCCACACAGAGAGAGGGG | 3960 |
| Dp | 3901 | GGTGTATCTTGGCACATATATCTCTGGGATGTAAGCTCACCCACACAGAGAGAGGGG | 3960 |
| OY | 3961 | AGAGTCAGGAGCTGTGTCTTACAGGCCCTATTAGGCCGTGACATCACCCCTTTCCTAGAAAT | 4020 |
| Dp | 3961 | AGAGTCAGGAGCTGTGTCTTACAGGCCCTATTAGGCCCTTGTACATCACCCCTTTCCTAGAAAT | 4020 |
| OY | 4021 | GGCCCCCTCATTTTTTGGTTCACATGATCTATTTTATACAGAGTGGGAGTGAAGCA | 4080 |
| Dp | 4021 | GGCCCCCTCATTTTTTGGTTCACATGATCTATTTTATACAGAGTGGGAGTGAAGCA | 4080 |
| OY | 4081 | AACTGCCACAGAAAGTTTGGGACTCCTCAGACCAGAAAGTATGTGCTCAGAAATCCCTGT | 4140 |
| Dp | 4081 | AACTGCCACAGAAAGTTTGGGACTCCTCAGACCAGAAAGTATGTGCTCAGAAATCCCTGT | 4140 |
| OY | 4141 | TCACCTTACAGTGGGGAATCTCCCTGGGGCTTCCAGGCTTGTTAGACAGAGGCT | 4200 |
| Dp | 4141 | TCACCTTACAGTGGGGAATCTCCCTGGGGCTTCCAGGCTTGTTAGACAGAGGCT | 4200 |
| OY | 4201 | ATTCCTTGTATAGGCGATGACACTAGTCTATGTGGTACTACATTCCTGTCCAGTTAAAG | 4260 |
| Dp | 4201 | ATTCCTTGTATAGGCGATGACACTAGTCTATGTGGTACTACATTCCTGTCCAGTTAAAG | 4260 |
| OY | 4261 | CTGGAACATAAAACCCACGCGCCAGGCTTCTCTACAGTTGTACCCCAAGAACACA | 4320 |
| Dp | 4261 | CTGGAACATAAAACCCACGCGCCAGGCTTCTCTACAGTTGTACCCCAAGAACACA | 4320 |
| OY | 4321 | AGACAGTAGATATGCAAGGATAGTATGCTGGGGAAGAAAGACTTAAACCCCCCAAG | 4380 |
| Dp | 4321 | AGACAGTAGATATGCAAGGATAGTATGCTGGGGAAGAAAGACTTAAACCCCCCAAG | 4380 |
| OY | 4381 | GCCACAGGTTCCGTTCCCTAGTTTCAACATGCGAGATAGGCTAGCTATATGAGGCTG | 4440 |
| Dp | 4381 | GCCACAGGTTCCGTTCCCTAGTTTCAACATGCGAGATAGGCTAGCTATATGAGGCTG | 4440 |
| OY | 4441 | TGAGTTGGTAGTACACAGCATGAGTATGTTGAGGCCACGCTTGAGCTATGAGGCA | 4500 |
| Dp | 4441 | TGAGTTGGTAGTACACAGCATGAGTATGTTGAGGCCACGCTTGAGCTATGAGGCA | 4500 |
| OY | 4501 | TTGGGAGGCTGAAGCAGAGGATTTGCTATATGTGGAAGCCACGCTTGAGCTATGAGGCA | 4560 |
| Dp | 4501 | TTGGGAGGCTGAAGCAGAGGATTTGCTATATGTGGAAGCCACGCTTGAGCTATGAGGCA | 4560 |
| OY | 4561 | GACTTGTCTTTAAGAAAAAATGAAGGCCACGACAGTGTGGACACGCTTTTAAATCCA | 4620 |
| Dp | 4561 | GACTTGTCTTTAAGAAAAAATGAAGGCCACGACAGTGTGGACACGCTTTTAAATCCA | 4620 |
| OY | 4621 | GCACTTGGGAGGAGAGGAGCATTTCTAGTTTCAAGGCGACGCTGTGCTATAGAGT | 4680 |
| Dp | 4621 | GCACTTGGGAGGAGAGGAGCATTTCTAGTTTCAAGGCGACGCTGTGCTATAGAGT | 4680 |
| OY | 4681 | GAGTTCCAGACAGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAACACAGAAAAACAA | 4740 |
| Dp | 4681 | GAGTTCCAGACAGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAACACAGAAAAACAA | 4740 |
| OY | 4741 | AACAAAAACAAACAAACAAACCCCAAAACCCCAAAACCCCAACCTCTCATCTCTCTCTC | 4800 |
| Dp | 4741 | AACAAAAACAAACAAACAAACCCCAAAACCCCAAAACCCCAACCTCTCATCTCTCTCTC | 4800 |
| OY | 4801 | TAGAGCTGTGCTGTAGTGTAGTGTGGGACTTCAGACTTATATTAATAGGCC | 4860 |

| | | | | |
|---|---|------|---|------|
| D | b | 4801 | TAGCTGTCGTCTAGTGTAGAGTTTGGGACTTCAGACTATATATATAATAGGC | 4860 |
| O | y | 4861 | TTTTTATCACTGTGTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCTTGAGA | 4920 |
| D | b | 4861 | TTTTTATCACTGTGTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGACCTTGAGA | 4920 |
| O | y | 4921 | AAGTCTCTCTTCCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT | 4980 |
| D | b | 4921 | AAGTCTCTCTTCCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT | 4980 |
| O | y | 4981 | CAGACTACTGTTCTGAAGGCGAGAGAGAGGTTTGAAGAAATGTGTGACAGACAGTT | 5040 |
| D | b | 4981 | CAGACTACTGTTCTGAAGGCGAGAGAGAGGTTTGAAGAAATGTGTGACAGACAGTT | 5040 |
| O | y | 5041 | GGAACAGAGGACAGAGAGGGGAGGACATCCAAAGATTCTGAACATGTAGCTGACTTTGGT | 5100 |
| D | b | 5041 | GGAACAGAGGACAGAGAGGGGAGGACATCCAAAGATTCTGAACATGTAGCTGACTTTGGT | 5100 |
| O | y | 5101 | TCTCTGGGTGACAGTGTCTCCCGAGGATAGGGCTGTAGAAAGGGGACCGAGGGTGAGCC | 5160 |
| D | b | 5101 | TCTCTGGGTGACAGTGTCTCCCGAGGATAGGGCTGTAGAAAGGGGACCGAGGGTGAGCC | 5160 |
| O | y | 5161 | AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG | 5220 |
| D | b | 5161 | AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG | 5220 |
| O | y | 5221 | AGAGCCCTCTAACCTCCCTGGAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTTCTA | 5280 |
| D | b | 5221 | AGAGCCCTCTAACCTCCCTGGAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTTCTA | 5280 |
| O | y | 5281 | GGGTGAAGAGAGATGTCGTCTGACCAACATGGCTAGGACAGAGAGATTTGGACCG | 5340 |
| D | b | 5281 | GGGTGAAGAGAGATGTCGTCTGACCAACATGGCTAGGACAGAGAGATTTGGACCG | 5340 |
| O | y | 5341 | TTACCCCTCAGAACCCGATCCCTCTTGGCTTAAGAGGCTGGGCCCTTTCTGTTT | 5400 |
| D | b | 5341 | TTACCCCTCAGAACCCGATCCCTCTTGGCTTAAGAGGCTGGGCCCTTTCTGTTT | 5400 |
| O | y | 5401 | AAGATCTTACTTCTTCTCTAGAGAGAGGACGACCTTTGTCCTCTCTTGTGTCAA | 5460 |
| D | b | 5401 | AAGATCTTACTTCTTCTCTAGAGAGAGGACGACCTTTGTCCTCTCTTGTGTCAA | 5460 |
| O | y | 5461 | TAAACACCCCTGTGTGAACATTAGTTTATTACTGTGACAGTTTCTCCAGAGAGTCCA | 5520 |
| D | b | 5461 | TAAACACCCCTGTGTGAACATTAGTTTATTACTGTGACAGTTTCTCCAGAGAGTCCA | 5520 |
| O | y | 5521 | TCTGTAGACCTCTGCTCTTAACCAAGGTATGGCCACATTCCTCACCCAGAAAG | 5580 |
| D | b | 5521 | TCTGTAGACCTCTGCTCTTAACCAAGGTATGGCCACATTCCTCACCCAGAAAG | 5580 |
| O | y | 5581 | TGCAGAAAGAGCCTTAGAGAAAGGTTAAACAGTAAACAAAGATGGCCAGATTAACAA | 5640 |
| D | b | 5581 | TGCAGAAAGAGCCTTAGAGAAAGGTTAAACAGTAAACAAAGATGGCCAGATTAACAA | 5640 |
| O | y | 5641 | ACTACTATCTTGTGACCAAAATTTGTTTGTCTGAACAGAGAGGGGTGTGAGTGTAT | 5700 |
| D | b | 5641 | ACTACTATCTTGTGACCAAAATTTGTTTGTCTGAACAGAGAGGGGTGTGAGTGTAT | 5700 |
| O | y | 5701 | GT | 5760 |
| D | b | 5701 | GT | 5760 |
| O | y | 5761 | CTTGGGGACCTTTCATGCTAAAGAAATATCTGATATTGGCCGCCATGCCAACAGGGGTAT | 5820 |
| D | b | 5761 | CTTGGGGACCTTTCATGCTAAAGAAATATCTGATATTGGCCGCCATGCCAACAGGGGTAT | 5820 |
| O | y | 5821 | TGGGAGAGTCAAGCTTGTGCAACACAGTATGCTCCCAAGATGGATTGGGGCTGAA | 5880 |
| D | b | 5821 | TGGGAGAGTCAAGCTTGTGCAACACAGTATGCTCCCAAGATGGATTGGGGCTGAA | 5880 |
| O | y | 5881 | TCACCAAGGGGAGGCTATCAGATGAGACAGAAATCAAGATTAAGCACTGTGGG | 5940 |
| D | b | 5881 | TCACCAAGGGGAGGCTATCAGATGAGACAGAAATCAAGATTAAGCACTGTGGG | 5940 |
| D | b | 5881 | TCACCAAGGGGAGGCTATCAGATGAGACAGAAATCAAGATTAAGCACTGTGGG | 5940 |
| O | y | 5941 | GCTCAGAAAGGAGTATTACAAGAGTTAAAGGCCAAGCCATTATTATTCAGAGATAGC | 6000 |
| D | b | 5941 | GCTCAGAAAGGAGTATTACAAGAGTTAAAGGCCAAGCCATTATTATTCAGAGATAGC | 6000 |
| O | y | 6001 | TCAAAATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGGCTGTGAGTGGGACACC | 6060 |
| D | b | 6001 | TCAAAATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGGCTGTGAGTGGGACACC | 6060 |
| O | y | 6061 | TGACCTTGCATTTATTAGTCACTAGGCCAAGAGGACATCACAGAGGTGACTGGTCTTA | 6120 |
| D | b | 6061 | TGACCTTGCATTTATTAGTCACTAGGCCAAGAGGACATCACAGAGGTGACTGGTCTTA | 6120 |
| O | y | 6121 | CTCAGCTTGGAGCAGGACAGCTGTGCAATGGTGACCTCATCTGATGAGAGGGCTGAG | 6180 |
| D | b | 6121 | CTCAGCTTGGAGCAGGACAGCTGTGCAATGGTGACCTCATCTGATGAGAGGGCTGAG | 6180 |
| O | y | 6181 | CACACAGGTTCAAGTGTCTCCCTGTGTCTCATATGCCAGGATTCCTGGCCAGTTTCAAG | 6240 |
| D | b | 6181 | CACACAGGTTCAAGTGTCTCCCTGTGTCTCATATGCCAGGATTCCTGGCCAGTTTCAAG | 6240 |
| O | y | 6241 | GACTAAGACATCTCTGTGTGGAACAAAGTATCCAAAGCCCTAAGCCCATTTTGTCT | 6300 |
| D | b | 6241 | GACTAAGACATCTCTGTGTGGAACAAAGTATCCAAAGCCCTAAGCCCATTTTGTCT | 6300 |
| O | y | 6301 | AATTAATTCAGAACCCCTGGGGATGCAAGCTGTGAGCAGACAGACTTTTAAAAAGCTC | 6360 |
| D | b | 6301 | AATTAATTCAGAACCCCTGGGGATGCAAGCTGTGAGCAGACAGACTTTTAAAAAGCTC | 6360 |
| O | y | 6361 | CCAGGTATTCGATCAGCAGCTGGAACAAACACAGCTACAGTTTCAAAACAGAAAGAGC | 6420 |
| D | b | 6361 | CCAGGTATTCGATCAGCAGCTGGAACAAACACAGCTACAGTTTCAAAACAGAAAGAGC | 6420 |
| O | y | 6421 | AAAGCTAAGGAAAGCTTGGGATGGGAGCTTCTTCAGGCCAGTAGATGAGGCTGTT | 6480 |
| D | b | 6421 | AAAGCTAAGGAAAGCTTGGGATGGGAGCTTCTTCAGGCCAGTAGATGAGGCTGTT | 6480 |
| O | y | 6481 | AGCAGTGTGGAGGCTCTCTGCTGCTGATATAGCTATTCATCCACATCATCCATCAT | 6540 |
| D | b | 6481 | AGCAGTGTGGAGGCTCTCTGCTGCTGATATAGCTATTCATCCACATCATCCATCAT | 6540 |
| O | y | 6541 | ACACCCACCCATCATTTATGACCAACCTTCATCCATCATCATCATCATCATCATCAT | 6600 |
| D | b | 6541 | ACACCCACCCATCATTTATGACCAACCTTCATCCATCATCATCATCATCATCATCAT | 6600 |
| O | y | 6601 | CCAGCATCATCATCAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 6660 |
| D | b | 6601 | CCAGCATCATCATCAACCTTCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 6660 |
| O | y | 6661 | CATTATCCACAGAGACTGTGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG | 6720 |
| D | b | 6661 | CATTATCCACAGAGACTGTGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG | 6720 |
| O | y | 6721 | AGCTCTGTTGATTCAGTATTTGTCATGTATGAGACAGGTATCATACACAGACAGCTG | 6780 |
| D | b | 6721 | AGCTCTGTTGATTCAGTATTTGTCATGTATGAGACAGGTATCATACACAGACAGCTG | 6780 |
| O | y | 6781 | TGGCAATGGAGAAAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG | 6840 |
| D | b | 6781 | TGGCAATGGAGAAAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG | 6840 |
| O | y | 6841 | TTGAACCTCAAAATTTATCGGGCTGGGCAAGTGTCTTTACCAACCGAGCCATTTTGTGACA | 6899 |
| D | b | 6841 | TTGAACCTCAAAATTTATCGGGCTGGGCAAGTGTCTTTACCAACCGAGCCATTTTGTGACA | 6899 |
| O | y | 6901 | CATCATTTATTTATTAAGAGCATCTATGATGACAGGCTGAGGCTCAAGCTTCTATGTCG | 6960 |
| D | b | 6901 | CATCATTTATTTATTAAGAGCATCTATGATGACAGGCTGAGGCTCAAGCTTCTATGTCG | 6960 |
| O | y | 6961 | CCACGGATGACCTTTTAACTCTGCTCTTCACAGCTTCACCCGAGGTAGGTTTACAGGT | 7020 |
| D | b | 6961 | CCACGGATGACCTTTTAACTCTGCTCTTCACAGCTTCACCCGAGGTAGGTTTACAGGT | 7020 |

| | | | |
|----|------|--|------|
| QY | 7021 | GTTAACTGGGTAATGGCTTTAATCCAGACACTGTGGGGGGGGGGGAGGAGGATC | 7088 |
| Db | 7020 | GTTCAACAGGGGAATGGCTTTAATCCAGACACTGTGGGGGGGGGGGAGGAGGATC | 7079 |
| QY | 7081 | CCTGAGTTGGAGGCGCACTTTGGTCTACAGAGTTTCAGCATACCTGGGCGTTACAGGAA | 7140 |
| Db | 7080 | CCTGAGTTGGAGGCGCACTTTGGTCTACAGAGTTTCAGCATACCTGGGCGTTACAGGAA | 7139 |
| QY | 7141 | ACCTATCCCAACAACAACAACAATAAATATTCTGTCAATATACACAGA | 7200 |
| Db | 7140 | ACCTATCCCAACAACAACAACAATAAATATTCTGTCAATATACACAGA | 7199 |
| QY | 7201 | TTAGAGGATTAATTAGTAGAGTAGAGGGCGTAGAGGAGAGCATGCTTTCTTTGTATT | 7260 |
| Db | 7200 | TTAGAGGATTAATTAGTAGAGTAGAGGGCGTAGAGGAGAGCATGCTTTCTTTGTATT | 7259 |
| QY | 7261 | ATAATAGTAAGTACTCAAGAAGTGCATTATCTATCTATCTATCTATCTATCTATC | 7320 |
| Db | 7260 | ATAATAGTAAGTACTCAAGAAGTGCATTATCTATCTATCTATCTATCTATCTATC | 7319 |
| QY | 7321 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC | 7380 |
| Db | 7320 | TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC | 7379 |
| QY | 7381 | CTGCTTTGACCTCAATAGCTCCTATTTCGGGGCAACCTTACCCCTAGTGGGGTTT | 7440 |
| Db | 7380 | CTGCTTTGACCTCAATAGCTCCTATTTCGGGGCAACCTTACCCCTAGTGGGGTTT | 7439 |
| QY | 7441 | ACCAACACCCAGACATTATTATTGTTTTATTATTAATCTATCTAGAGCTCAGGGTG | 7500 |
| Db | 7440 | ACCAACACCCAGACATTATTATTGTTTTATTATTAATCTATCTAGAGCTCAGGGTG | 7499 |
| QY | 7501 | GGACATCAGGGCTGTGTGCATGCTAATACCAAGCTCTGCCACAGACCTCAGTCCAGTCC | 7560 |
| Db | 7500 | GGACATCAGGGCTGTGTGCATGCTAATACCAAGCTCTGCCACAGACCTCAGTCCAGTCC | 7559 |
| QY | 7561 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCAATATTCGACGGCTATGTAGCTCTTC | 7620 |
| Db | 7560 | CCATTTTGTTCAGGTGACTGTGTGACAGTTGTCAATATTCGACGGCTATGTAGCTCTTC | 7619 |
| QY | 7621 | CACGCCAGTCCAGCATCTTCTGTGCATCCAGTGGGGGGGCAACCTGGTGCACACA | 7680 |
| Db | 7620 | CACGCCAGTCCAGCATCTTCTGTGCATCCAGTGGGGGGGCAACCTGTGCTGCACACA | 7679 |
| QY | 7681 | GTCGCCGTGTTCCCTGTCTTCAGACACTACATATTTCGCTGTGGAACAGTTGATGAATG | 7740 |
| Db | 7680 | GTCGCCGTGTTCCCTGTCTTCAGACACTACATATTTCGCTGTGGAACAGTTGATGAATG | 7739 |
| QY | 7741 | GGATGCGCTTCTGTGTAATCTTTTAATGGCTGGGCCCTTATCTTAGCACAGTTTGTGTG | 7800 |
| Db | 7740 | GGATGCGCTTCTGTGTAATCTTTTAATGGCTGGGCCCTTATCTTAGCACAGTTTGTGTG | 7799 |
| QY | 7801 | GGCCATGTGCACGCTATCTATCTATCTATCATCATCTTAATAGCGTTAATAGTGTCTT | 7860 |
| Db | 7800 | GGCCATGTGCACGCTATCTATCTATCTATCATCATCTTAATAGCGTTAATAGTGTCTT | 7859 |
| QY | 7861 | TGTGTGATTAACACACTTCTGTTCATTACGATGGAATAATTTGTGGCCCCACCCAC | 7920 |
| Db | 7860 | TGTGTGATTAACACACTTCTGTTCATTACGATGGAATAATTTGTGGCCCCACCCAC | 7919 |
| QY | 7921 | CCTTTTTTTTTTTTATTTGAACAAGGCTTTTCTGTGTATCTTGGCAATCTTGGCTGTCT | 7980 |
| Db | 7920 | CCTTTTTTTTTTTTATTTGAACAAGGCTTTTCTGTGTATCTTGGCAATCTTGGCTGTCT | 7979 |
| QY | 7981 | GAGCTCACTGTGTAGACAGGCGTGTAGGCTGTCTTCACACTTTTGACACCTCCTGTGAC | 8040 |
| Db | 7980 | GAGCTCACTGTGTAGACAGGCGTGTAGGCTGTCTTCACACTTTTGACACCTCCTGTGAC | 8039 |
| QY | 8041 | AGAGTACCATGAACTTCAAAAGCAATTTTCTGTGTTTGGTTTGTATTTACATTTGTGTG | 8100 |
| Db | 8040 | AGAGTACCATGAACTTCAAAAGCA - TTTCTGTGTTTGGTTTGTATTTACATTTGTGTG | 8098 |

| | | | |
|----|-------|--|-------|
| OY | 810.1 | TGATGGGTGATATATGTCATGTTTGGTTCAGGTGCTGCACATGTCATACCGTGTG | 816.0 |
| Db | 809.9 | TGTAATGGGTGATATATGTCATGTTTGGTTCAGGTGCTGCACATGTCATACCGTGTG | 815.8 |
| OY | 816.1 | TGGACACAGAAACAACCCGATGTGCCATTCCTCAGATCTACTACGCATCTTGTTAAATGTA | 822.0 |
| Db | 815.9 | TGGGACAGAGAAACAACCCGATGTGCCATTCCTCAGATCTACTACGCATCTTGTTAAATGTA | 821.8 |
| OY | 822.1 | TGTAATATGATGTTTATTTAGTGTGCTCCCAAGTATGCAAGTAATTTGTTGGAGTTTTCAC | 828.0 |
| Db | 821.9 | TGTAATATGATGTTTATTTAGTGTGCTCCCAAGTATGCAAGTAATTTGTTGGAGTTTTCAC | 827.8 |
| OY | 828.1 | CTTGCCCTTGAGGGCTGTCGCCCATTTAAACATCAGTCTCGGGGTAGTGAAGCATGCTTCCA | 834.0 |
| Db | 827.9 | CTTGCCCTTGAGGGCTCTCCGCATTTAAACATCAGTCTCGGGGTAGTGAAGCATGCTTCCA | 833.8 |
| OY | 834.1 | CTCGATGAGCCATCTCGCTGCCCTCCCTGTCACACTCTCTCTTATTTTCCAGATGGGACTA | 840.0 |
| Db | 833.9 | CTCGATGAGCCATCTCGCTGCCCTCCCTGTCACACTCTCTCTTATTTTCCAGATGGGACTA | 839.8 |
| OY | 840.1 | CGCAGCTCAGCGCTTAAACCTCAGCAATCACAAGATCCAGATGGGCTAGCCAGGACATCTAG | 846.0 |
| Db | 839.9 | CGCAGCTCAGCGCTTAAACCTCAGCAATCACAAGATCCAGATGGGCTAGCCAGGACATCTAG | 845.8 |
| OY | 846.1 | GGATATGCTGGCCCTCTGCTCCACACAGTGTAGAAATTTACAGGCGATACATCAGCTGCTGAG | 852.0 |
| Db | 845.9 | GGATATGCTGGCCCTGCTGCTCCACACAGTGTAGAAATTTACAGGCGATACATCAGCTGCTGAG | 851.8 |
| OY | 852.1 | ATTTTTTAACCTGAATCTCTAGGATTAAGACAGGCACTCTACCAATGGAGGGTTCTTTTGTCT | 858.0 |
| Db | 851.9 | ATTTTTTAACCTGAATCTCTAGGATTAAGACAGGCACTCTACCAATGGAGGGTTCTTTTGTCT | 857.8 |
| OY | 858.1 | GTTTGGTTTGTTTCCCTCCTCATTAAGATCAGGCAAGTCTCAATATAGTATAGCTGGGCTAC | 864.0 |
| Db | 857.9 | GTTTGGTTTGTTTCCCTCCTCATTAAGATCAGGCAAGTCTCAATATAGTATAGCTGGGCTAC | 863.8 |
| OY | 864.1 | ATAACATCTTGCTCAAAAAGCCCTATTAAGAGTAGGGAGGCTGAGGCTTAAGAAGAGCCTT | 870.0 |
| Db | 863.9 | ATAACATCTTGCTCAAAAAGCCCTATTAAGAGTAGGGAGGCTGAGGCTTAAGAAGAGCCTT | 869.8 |
| OY | 870.1 | AAGCGCGCTGATAGACACACAGGATAGGCTGAGCTATATACCAGAGCACTTGTTTCAAAA | 876.0 |
| Db | 869.9 | AAGCGCGCTGATAGACACACAGGATAGGCTGAGCTATATACCAGAGCACTTGTTTCAAAA | 875.8 |
| OY | 876.1 | ACATGAGGAGGAGGGATGTGTTTTAAGTGTGGGCTGTGTAAACAGGCACTTAAGGAGGCCAA | 882.0 |
| Db | 875.9 | ACATGAGGAGGAGGGATGTGTTTTAAGTGTGGGCTGTGTAAACAGGCACTTAAGGAGGCCAA | 881.8 |
| OY | 882.1 | TGTAAGCATTTTGACTAAGAAAGGATCATCAACCCGGGTGGGACAGGTATAGAGTTGG | 888.0 |
| Db | 881.9 | TGTAAGCATTTTGACTAAGAAAGGATCATCAACCCGGGTGGGACAGGTATAGAGTTGG | 887.8 |
| OY | 888.1 | ACTACAGCTGTCAAGACCCCATAGGAGAGCCAGTTTCCCTTCTCTGAGGCTTCAAGC | 894.0 |
| Db | 887.9 | ACTACAGCTGTCAAGACCCCATAGGAGAGCCAGTTTCCCTTCTCTGAGGCTTCAAGC | 893.8 |
| OY | 894.1 | CTGGCTGTGACGGCCACTGCTCTCTACATGCCCTTCTCTTAGGCTGTGCACCATAG | 899.5 |
| Db | 893.9 | CTGGCTGTGACGGCCACTGCTCTCTACATGCCCTTCTCTTAGGCTGTGCACCATAG | 899.3 |

| | |
|-----------|---|
| AAAF25499 | RESULT 3 |
| ID | AAF25499 standard; DNA: 23024 BP. |
| XX | |
| AC | AAF25499; |
| XX | |
| DT | 15-MAY-2001 (first entry) |
| XX | |
| DE | Nucleotide sequence of a murine ABCA7 (ABCA7) transporter |
| XX | |
| XX | ABCA transporter; high-density lipoprotein cholesterol; |
| KW | ABCA7 transporter; ss. |
| FM | ABCA7 transporter; ss. |


```

XX Mus musculus.
OS
XX WO200109314-A1.
XX
XX 08-FEB-2001.
XX
XX 28-JAN-2000; 2000WO-FR00209.
XX
XX 30-JUL-1999; 99FR-0009926.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Chimini G;
XX
XX WPI; 2001-182953/18.
XX
XX Selecting agents that modulate ABCA transporters, useful e.g. for
XX normalizing serum cholesterol levels, comprises using transgenic
XX animals with an inactive ABCA gene allele.
XX
XX Claim 12; Fig 22A; 113pp; French.
XX
XX The specification describes a method for selecting or screening agents
XX that modulate ABCA transporters. The method comprises using non-human
XX recombinant mammals with an inactivated allele of the gene encoding the
XX ABCA transporter, or cells with an inactivated allele of ABCA
XX transporter, from any tissue of the recombinant mammal, preferably with
XX an allele truncated in one or both exons corresponding to the first
XX and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate
XX ABCA transporters may be useful for increasing (normalizing) serum
XX levels of high-density lipoprotein cholesterol. The present sequence
XX represents the genomic sequence of a murine ABCA7 (ABCA7) transporter.
XX
XX Sequence 23024 BP; 5195 A; 6205 C; 6348 G; 5266 T; 10 other:
XX
XX Query Match 1.7%; Score 153.2; DB 22; Length 23024;
XX Best Local Similarity 79.1%; Pred. No. 1.7e-26;
XX Matches 182; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
XX
XX 4552 ATAGACGAGACTTGTCTTAAGAAAAAATGAAAGCCAGCAGTGTGACAGCGCT 4611
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 16839 ACACAGAAACCCGTCTAAACAAACAAACAGCGGCGACGTGTGACAGCGCT 16898
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4612 TTAATCCAGCACTTGGGAGGAGCAAGACGAGTTCAGTTCAGGCGACCTGCT 4671
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 16899 TTAATCCAGCACTTGGGAGGAGGAGGAGGAGTTCAGTTCAGGCGTACCTGCT 16958
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4672 CTATAGAGTGTCCAGGACAGCGAGGCTACACAGAAACCCGTTTGAACAAACA 4731
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 16959 CTACATAGAGTGTCCAGGACAGCGAGCTACACAAACCCGTTTGAACAAACA 17018
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4732 GAAAAACAAACAAACAAACAAACAAACCAACCAACCAACCAACCAAC 4781
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 17019 AAAAAACAAACCAACCTTAACCAACCAACCAACCAACCAACCAACCAAC 17068
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX RESULT 4
XX AAL38337
XX ID AAL38337 standard; DNA; 215980 BP.
XX
XX AAL38337;
XX
XX 15-AUG-2002 (first entry)
XX
XX Complementary strand of a genomic sequence encoding a mouse Ngr3.
XX
XX Cerebroprotective; neuroprotective; cytosstatic; Nogo receptor homologue;
XX Ngr3; axonal growth; central nervous system; CNS; cerebral injury;
XX spinal cord injury; stroke; demyelinating disease; multiple sclerosis;
XX monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;
XX multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;

```

```

XX Alexander's disease; Canavan's disease; metachromatic leukodystrophy;
XX Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;
XX transgenic animal; unregulated cellular growth; cancer; tumour; mouse;
XX murine; ds.
XX
XX Mus sp.
XX
XX WO200229059-A2.
XX
XX 11-APR-2002.
XX
XX 06-OCT-2001; 2001WO-US31488.
XX
XX 06-OCT-2000; 2000US-238361P.
XX
XX (UYTA ) UNIV YALE.
XX (BIOJ ) BIOGEN INC.
XX
XX Stittmatter SM, Cate RL, Sah DWY;
XX
XX WPI; 2002-416677/44.
XX
XX Novel Nogo receptor homolog polypeptide, Ngr2 or Ngr3, useful for
XX treating central nervous system disorder, cerebral injury, spinal cord
XX injury, stroke, and demyelinating diseases.
XX
XX Example 16; Page 219-275; 277pp; English.
XX
XX The invention relates to a Nogo receptor homologue polypeptide, Ngr2 or
XX Ngr3, comprising a 50 amino acid LRRC2 sequence, a 284 amino acid NTLRRC2
XX sequence, or a 420, 461 or 392 amino acid sequence, all given in the
XX specification. The Ngr3 protein or its binding antibody is useful for
XX decreasing inhibition of axonal growth of a central nervous system (CNS)
XX neuron, by contacting the neuron Ngr3 or its antibody, and for treating
XX CNS disease, disorder or injury. Ngr3 or a vector comprising Ngr3 is
XX useful for treating cerebral injury, spinal cord injury, stroke,
XX demyelinating diseases, e.g. multiple sclerosis, monophasic
XX demyelination, encephalomyelitis, multifocal leukoencephalopathy,
XX panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,
XX Alexander's disease, Canavan's disease, metachromatic leukodystrophy and
XX Krabbe's disease. Ngr3 is useful for inducing an immune response in a
XX mammal against Ngr3, as a bait protein in a two-hybrid or three-hybrid
XX assay, and as a research tool for identification, characterisation and
XX purification of interacting, regulatory proteins. The nucleotide
XX sequences of the invention are useful for screening for RFLP associated
XX with certain disorders, for genetic mapping, and for gene therapy. The
XX vector containing Ngr3 is useful for producing non-human transgenic
XX animals. The Ngr3 binding antibody is useful for isolating and purifying
XX Ngr3, for localisation and/or quantitation of Ngr3, and for diagnostic
XX and therapeutic purposes. The sequences of the invention, vectors and
XX antibodies are useful for treating or preventing unregulated cellular
XX growth such as cancer and tumour growth. This polynucleotide sequence
XX represents the
XX
XX Sequence 215980 BP; 56872 A; 50995 C; 50240 G; 54437 T; 3436 other:
XX
XX Query Match 1.7%; Score 152.2; DB 24; Length 215980;
XX Best Local Similarity 82.2%; Pred. No. 9.6e-26;
XX Matches 175; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
XX
XX 4560 AGACATTTGCTTTAAGAAAAAATGAAAGCCAGCAGTGTGACAGCGCTTAATGCC 4619
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 119710 ATATATTATTACTACATTAAACAAACAGCGCGGTGTGCGCATCTTAATGCC 119769
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4620 AGCATTGGAGAGCAAGACAGATTTCTGAGTTCAAGCAGCGCTGTATAGAG 4679
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 119770 AGCATTGGAGAGCAAGACAGCGGATTTCTGAGTTCAAGCAGCGCTGTATAGAG 119829
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4680 TGAATTCCAGGACAGCGGCTACACAGAAACCTGTTTGAACCAACGAAAAACA 4739
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 119830 TTAGTTCCAGGACAGCGGCTACACAGAAACCTGTTTGAACCAACGAAAAACA 119889
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 4740 AAAAAACAAACAAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4772

```

Db 119890 AAAAAAAAAAAAAAAAAACCAACCAACACA 119922

RESULT 5

AAAC82696 AAC82696 standard; cDNA; 1445 BP.

AC AAC82696;

DT 15-MAR-2001 (first entry)

DE Murine variant zalpha32 cDNA.

KM Zalpha32; murine; secreted alpha-helical protein-32; antiinflammatory;

KM spermatogenesis; steroidogenesis; testicular differentiation;

KM regulatory control; hypothalamic-pituitary-gonadal axis;

KW reproductive system disorder; immunological system disorder; ss.

OS Mus musculus.

FH Key Location/Qualifiers

FT CDS 19..624

FT /tag= a /product= "variant Zalpha32"

PN MO200071717-A1.

PD 30-NOV-2000.

PE 25-MAY-2000; 2000MO-US14563.

PR 26-MAY-1999; 99US-0320159.

PA (ZYMO) ZYMOGENETICS INC.

PI Conklin DC, Gao Z;

DR WPI: 2001-032038/04.

DR P-PSDB; AAB45760.

PT Novel mammalian secreted alpha-helical protein-32 (Zalpa32) useful for

PT identifying antagonists of its activity that are used in treating

PT Zalpa32-induced inflammation

PS Disclosure: Page 62-63; 73pp; English.

CC This invention describes a novel mammalian secreted alpha-helical
CC protein-32 (Zalpa32) (I). The invention also describes (I) an isolated
CC polynucleotide (II) encoding (I); (2) an antibody (III) that specifically
CC binds to (I); and (3) an educational kit (IV) for the teaching of
CC molecular biology and/or biochemistry comprising (II). The products of
CC the invention have antiinflammatory activity. (III) as an antagonist to
CC Zalpa32 is useful for treating Zalpa32-induced inflammation. Zalpa32
CC antibodies that specifically bind to Zalpa32 epitopes. Antibodies or
CC polypeptides, directly or indirectly conjugated to drugs or toxins are
CC useful for in vivo diagnostic or therapeutic applications.
CC Zalpa32-cytokine fusion proteins are useful for enhancing in vivo
CC killing of target tissue. (I) and (II) are useful to identify and isolate
CC receptors involved in spermatogenesis, steroidogenesis, testicular
CC differentiation and regulatory control of the
CC hypothalamic-pituitary-gonadal axis and also for testing disorders of the
CC reproductive system and immunological systems.

SQ Sequence 1445 BP; 418 A; 319 C; 398 G; 310 T; 0 other;

Query Match 1.7%; Score 150.6; DB 22; Length 1445;

Best Local Similarity 83.4%; Pred. No. 1.8e-26;

Matches 111; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 4570 TTTAAGAAAAAATGAAGCCAGCACTGTGCGACACGCTTTAATCCAGCACTTGGG 4629

Db 988 TTTAAAAAATGTGTGCTTACCGCGCGTGTGCGCACGCTTTAATCCAGCACTTGGG 1047
OY 4630 AGCGAGAGAGCAGCAGATTCTGAGTTCAAGCCAGCCTGCTATAGATGATTCAG 4689
Db 1048 AGCGAGAGCAGCAGGATTTCTGAGTTCAAGCCAGCCTGCTATAGATGATTCAG 1107
OY 4690 GACAGCCAGGCTTACAGAGAAACCTGTTTGAAGAAACAGAAACCAACCAACA 4749
Db 1108 GACAGCCAGGCTTATACAGAGAAACCTGCTTGAAGAAACCAACCAACA 1167
OY 4750 AAACAAACCAAAACCCAAACCCAAA 4774
Db 1168 ACAACCAAAACCAAAACCAAA 1192

RESULT 6

AAA40866 AAA40866 standard; DNA; 7208 BP.

AC AAA40866;

DT 16-AUG-2000 (first entry)

DE Murine tumour necrosis factor alpha (TNFalpha) nucleotide sequence.

KM Antisense oligonucleotide; phosphorothioate; TNFalpha; cytokine; inhibit;

KM tumour necrosis factor alpha; inflammatory bowel disease; diabetes;

KM rheumatoid arthritis; infectious disease; multiple sclerosis; hepatitis;

KW pancreatitis; atopic dermatitis; allograft rejection;

KW autoimmune disease; inflammatory disease; ds.

OS Mus sp.

PN MO200020645-A1.

PD 13-APR-2000.

PE 05-OCT-1999; 99MO-US23205.

PR 05-OCT-1998; 98US-0166186.

PR 18-MAY-1999; 99US-0313932.

PA (ISIS-) ISIS PHARM INC.

PI Baker BF, Bennett CF, Butler MM, Shanahan WJ;

DR WPI: 2000-303808/26.

DR P-PSDB; AAY88596.

PT Oligonucleotide for treating diseases associated with human tumour
PT necrosis factor-alpha (TNFalpha) such as, diabetes and rheumatoid
PT arthritis, comprises nucleotide sequence complementary to intron of
PT nucleic acid encoding TNFalpha -

PS Example 8; Page 173-179; 283pp; English.

CC This sequence represents the murine tumour necrosis factor alpha
CC (TNFalpha) nucleotide sequence. TNFalpha is an important cytokine that
CC plays a role in host defence. It is produced mainly in macrophages and
CC monocytes in response to infection, invasion, injury or inflammation.
CC Overexpression of TNFalpha can result in disease states, particularly in
CC infectious, inflammatory and autoimmune diseases. The invention relates
CC to antisense oligonucleotides which are capable of modulating the
CC TNFalpha gene expression. The oligonucleotides optionally have a
CC phosphorothioate backbone, and may also optionally contain at least one
CC 2'-O-methoxyethyl modification. The oligonucleotides are useful for
CC modulating the expression of human TNFalpha in cells and tissues,
CC reducing a human cell inflammatory response, reducing the blood glucose
CC level in a human and treating a human having a disease or condition
CC associated with TNFalpha. Examples of diseases associated with TNFalpha
CC include diabetes, inflammatory bowel disease, multiple sclerosis,
CC pancreatitis, rheumatoid arthritis, infectious disease, hepatitis, atopic
CC dermatitis or allograft rejection. The antisense oligonucleotides are

CC also useful for modulating the function of a selected nucleic acid
CC sequence in adipose tissue.
XX
SQ Sequence 7208 BP; 1750 A; 2011 C; 1810 G; 1637 T; 0 other;
Query Match 1.7%; Score 148.8; DB 21; Length 7208;
Best Local Similarity 88.0%; Pred. No. 1.1e-25;
Matches 162; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 4574 AGAAAAAATGAACCCAGCAGTGTGGCACACCCCTTTAATCCAGACTTGGAGGC 4633
D 429 AAAAAAAAAAAAAAGCTGGCAGTGTGGCACACACCTTTAATCCAGACTTGGAGGC 488
QY 4634 AGAAGCAGCAGATTTCTGAGTTCAAGCCAGCCCTGTCTATAGAGTTCAGAGCA 4693
D 489 AGAGCAGCGGATTTCTGAGTTCTAAGCCAGCCCTGTCTACAGAGTTCAGAGCA 548
QY 4694 GCCAGGCTACACAGAGAACCCTGTTTGAACCAACCAACCAACCAACCAAC 4753
D 549 GCCAGGCTACACAGAGAACCCTGTTCTGCAACCAACCAACCAACCAACCAAC 608
QY 4754 AAAA 4757
D 609 AAAA 612
RESULT 7
AAF63436/C
ID AAF63436 standard; DNA; 6789 BP.
XX
AC AAF63436;
XX
DT 14-MAY-2001 (first entry)
XX
DE Murine CD39-L4 genomic DNA sequence.
XX
KW Human CD39-like protein; apyrase; NDase; platelet function inhibitor;
KW myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;
KW cerebral artery thrombosis; platelet aggregation; inflammation;
KW apoptosis; autoimmune disorder; neurological disorder;
KW Alzheimer's disease; Parkinson's disease; cancer; CD39-L2; ds.
XX
OS Mus sp.
XX
PN MO200110205-A1.
XX
PD 15-FEB-2001.
XX
PF 09-AUG-2000; 2000MO-US21790.
XX
PR 09-AUG-1999; 990US-0370265.
PR 11-JAN-2000; 2000US-0481238.
PR 25-APR-2000; 2000US-0557800.
PR 26-MAY-2000; 2000US-0583231.
PR 30-JUN-2000; 2000US-0608285.
XX
RA (HYSE-) HYSEQ INC.
XX
PI Ford J, Mulero JJ, Yeung G;
XX
DR WPI; 2001-147489/15.
XX
PT Polynucleotides encoding human CD39-like polypeptides, with apyrase
PT and/or NDase activity, which are useful in the treatment of
PT pathological conditions caused by thrombosis (e.g. myocardial
PT infarction) and inflammatory disorders -
XX
PS Example 21; Page 108-110; 203pp; English.
XX
CC This invention relates to polynucleotides encoding human CD39-like
CC polypeptides with apyrase and/or NDase activity. The polypeptides having
CC ATPase, including NDase, activity are useful for inhibiting platelet
CC function and can therefore be used in the prophylaxis or treatment of

CC pathological conditions caused by or involving thrombosis or excessive
CC coagulation or excessive platelet aggregation, such as myocardial
CC infarction, cerebral ischemia, angina, arterial thrombosis, cerebral
CC artery thrombosis or intracardiac thrombosis, and conditions associated
CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in
CC modulating disease states (including platelet aggregation, inflammation
CC and apoptosis) associated with ADP or other purinergic signaling by
CC reducing the levels of NDps. The polypeptides are also useful for
CC prophylaxis or treatment of inflammation related disorders, such as
CC disorders involving sepsis or systemic inflammatory response syndrome or
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,
CC cytokine overstimulation); autoimmune disorders such as thrombosis,
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;
CC neurological disorders including neurodegenerative diseases, epilepsy,
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, and amyotrophic lateral sclerosis; and cancer. The present
CC sequence represents the murine CD39 like protein CD39-L4 genomic DNA
CC sequence.
XX
SQ Sequence 6789 BP; 1775 A; 1476 C; 1582 G; 1952 T; 4 other;
Query Match 1.6%; Score 145.6; DB 22; Length 6789;
Best Local Similarity 62.3%; Pred. No. 6.8e-25;
Matches 304; Conservative 0; Mismatches 164; Indels 20; Gaps 4;
QY 1854 GGAGGAGAGATGGCTGACGTTCCAGAGCACCTCTCTTGGCAGAGCACTGATTC 1913
D 1483 GGCCAGAGAGATGACTCGAAGATTATGCCCTCTCTTACAGAGGACCTGATTC 1424
QY 1914 AGTTCCAGAGA--CTCATATGGTGGCTCAGAGCCATCTGTAATCCAGTCCAGAGGTT 1971
D 1423 AGTTCCAGAGATGCTCAAGAGGTAGTTGTAATCTCTATATATTCACATCCAGAGATC 1364
QY 1972 CCACACCCCTTCTGGCCCTCCACAGGACACATCATATAGTACAGACATACATCAGG 2031
D 1363 TGACACCCCTCATGCAAGTATGACCCACACACAAATTAATAAATATTCATTTG 1304
QY 2032 CAAACACCCATACACATATAATTAATAGGAACCTTAAGAGTGTGTTGTTAA 2091
D 1303 AAGAAACCTAAAGTATGATGATTCATTAATTAATGAACCAAGTCTACCTGGTTC 1244
QY 2092 CATGTGCTTACACATGCTGATTAAGACATGTACACAGCCACCTGAAGAGGATCTG 2151
D 1243 AAACATTTGTTTTTTTTTTTCTAGATGACACTAATAAATAAACAATAAG-----G 1190
QY 2152 GGGCTGAGAGATGGCTCAGCGGTTAAGAGCACTGCTCTTCCGAAGAGGTCCTG 2211
D 1189 GGGCTGTCAGATGGCTCAGAGGGTAAAGAGCACTGCTCTTCTTC-----GAAGTCTG 1134
QY 2212 AGTTCAATTCCTAGCAGACACATGCTGCTCAACATCCATTAATGAGATCTGACACC 2271
D 1133 AGTTCAATTCCTAGCAGACACATGCTGCTCAACATCCATTAATGAGATCTGACACC 1074
QY 2272 TCTTCTGTCATCTGAGAGACAGTGCAGAGCTACAGTACTTATGATTAATAAT 2331
D 1073 TCTTCTGTCATCTGAGAGTGC-----AGCTACAGTGTACTTATGATTAATAAT 1022
QY 2332 AAATCTTT 2339
D 1021 AAATCTTT 1014
RESULT 8
AAA47451/C
ID AAA47451 standard; CDNA; 2221 BP.
XX
AC AAA47451;
XX
DT 20-OCT-2000 (first entry)
XX
DE Mouse TANGO 232 coding sequence.
XX


```
Oy 4562 ACTTGTCTTTAAGAAAAATGAAAGCCACGACGTGGTGACACAGCCTTTAATCCCG 4621
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4668 ATTCTCTTTTAAAAAAGAAAAAGAGCCGCGATGGTGACACATGCCTTTAATCCCG 4727
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4632 CACTTGGAGCGAAGACGACGACATTTCTGATTCAGAGCCAGCCTGGTCTATAGAGTG 4681
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4728 CACTTGGAGCGAAGACGACGACATTTCTGATTCAGAGCCAGCCTGGATACAGAGTG 4787
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4662 AGTTCCAGACGACGACGCTTACAGAGAAACCTGTTTGAATAACCCAGAAAAACAA 4741
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4788 AGTTCCAGACGACGACGCTTACTCAGAGAAACCTGTTTGAATAACCCAGAAAAAGAG 4847
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4742 ACAAACAACAA 4754
    | | | | | | |
Db 4848 GAAAGAAAAAAGA 4860

RESULT 10
AAx23318
ID AAX23318 standard; DNA: 1698 BP.
AC AAX23318;
XX
XX 11-JUN-1999 (first entry)
DT
DE Mouse 1-alpha-OHase promoter region DNA #3.
XX
XX 1-alpha-OHase: vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW diagnosis; treatment; disorder; endocrine; promoter; ss.
XX
XX Mus sp.
OS
XX WO9907835-A2.
PN
XX 18-FEB-1999.
PD
XX 06-AUG-1998; 98WO-CA00758.
PF
XX 06-AUG-1997; 97US-0906791.
PR
XX (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PA
XX PI Glorieux FH, St-Arnaud R;
XX
XX WPI: 1999-190048/16.
DR
XX
XX New polynucleotides encoding 25-hydroxyvitamin D 1- $\alpha$ -hydroxylase -
PT useful to detect and treat vitamin D-related disorders
PS
XX Claim 12; Fig 15; 85pp; English.
XX
XX This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC pathway, and polynucleotides encoding the enzyme from human, mouse and
CC rat tissue. Molecules of the invention are used to diagnose and treat
CC vitamin D-related disorders, to produce vitamin D metabolites, and to
CC identify modulators of the enzyme expression. The invention describes
CC an animal cell transformed with the enzyme promoter which is used to
CC identify compounds which modulate activity of the polynucleotide
CC encoding the enzyme, and to modulate the production of vitamin D in
CC patients with vitamin D endocrine disorders in the vitamin D
CC hydroxylation pathway, the enzyme catalyses conversion of
CC 25-hydroxyvitamin D to 1 alpha 25(OH)2D3.
XX
XX
XX Sequence 1698 BP; 499 A; 425 C; 414 G; 360 T; 0 other;
SQ
Query Match 1.6%; Score 141.4; DB 20; Length 1698;
Best Local Similarity 77.1%; Pred. No. 3.6e-24;
Matches 172; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 4552 ATAGAGCAGACTTGTCTTTAAGAAAAATGAAGCCAGCAGTGGTGACACAGCCT 4611
```

```
Db 923 AAAAGGCAAAATTTGTCATATTAACAGTGAATCTAAGCCGGGGCGTGGTGACAGCCT 982
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4612 TTAATCCAGACACTTGGAGCGACAGACGACGATTTTGTGATTCAGCCAGCCTGGT 4671
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 983 TTAATCCAGACACTTGGAGCGACAGACGACGCTGATTTCTGATTCATGCGACGCTGGT 1042
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4672 CTATGAGTGAAGTTCAGACGACGCGCTTACAGAGAAACCTGTTTGAATAACCA 4731
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1043 CTACAGAGTGAAGTTCAGACGACGCGCTTACAGAGAAACCTGTTTGGGAAAAACA 1102
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 4732 GAAAAACAACCAAAACAACCAAAACCAAAACCAAA 4774
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1103 AAAAAACAACCAAAACAACCAAAACCAAAACCAAGTACTTAA 1145

RESULT 11
AAx23316
ID AAX23316 standard; DNA: 4105 BP.
AC AAX23316;
XX
XX 11-JUN-1999 (first entry)
DT
DE Mouse 1-alpha-OHase promoter region DNA #2.
XX
XX 1-alpha-OHase: vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW diagnosis; treatment; disorder; endocrine; promoter; ss.
XX
XX Mus sp.
OS
XX WO9907835-A2.
PN
XX 18-FEB-1999.
PD
XX 06-AUG-1998; 98WO-CA00758.
PF
XX 06-AUG-1997; 97US-0906791.
PR
XX (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PA
XX PI Glorieux FH, St-Arnaud R;
XX
XX WPI: 1999-190048/16.
DR
XX
XX New polynucleotides encoding 25-hydroxyvitamin D 1- $\alpha$ -hydroxylase -
PT useful to detect and treat vitamin D-related disorders
PS
XX Example 4; Page 79-81; 85pp; English.
XX
XX This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC pathway, and polynucleotides encoding the enzyme from human, mouse and
CC rat tissue. Molecules of the invention are used to diagnose and treat
CC vitamin D-related disorders, to produce vitamin D metabolites, and to
CC identify modulators of the enzyme expression. The invention describes
CC an animal cell transformed with the enzyme promoter which is used to
CC identify compounds which modulate activity of the polynucleotide
CC encoding the enzyme, and to modulate the production of vitamin D in
CC patients with vitamin D endocrine disorders in the vitamin D
CC hydroxylation pathway, the enzyme catalyses conversion of
CC 25-hydroxyvitamin D to 1 alpha 25(OH)2D3.
XX
XX
XX Sequence 4105 BP; 1063 A; 1032 C; 1063 G; 947 T; 0 other;
SQ
Query Match 1.6%; Score 141.4; DB 20; Length 4105;
Best Local Similarity 77.1%; Pred. No. 5.7e-24;
Matches 172; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Oy 4552 ATAGAGCAGACTTGTCTTTAAGAAAAATGAAGCCAGCAGTGGTGACACAGCCT 4611
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3335 AAAAGGCAAAATTTGTCATATTAACAGTGAATCTAAGCCGGGGCGTGGTGACAGCCT 3394
```

OY 4612 TTATCCACGACCTTGGAGGAGCAAGACGAGATTTCGAGTTCAAGGCCAGCCTGGT 4671
|||||
DB 3395 TTATATCCACGACCTTGGAGGAGGACGAGCTGATTTCTGAGTTCATGGCCAGCCTGGT 3454
OY 4672 CTATAGAGTAGTATGCCAGGACGAGGCTATACAGAGAAACCCGTGTTGAAAAACA 4731
|||||
DB 3455 CTACAGAGTAGTATGCCAGGACGAGGCTTACAGAGAAACCCGTGTTGAAAAACA 3514
OY 4732 GAAAAACAAACAAACAAACAAACAAACCCCAAA 4774
|||||
DB 3515 AAAAACAAACAAACAAACAAACAAACCCCAAGTACTCTAA 3557

RESULT 12
AAL47852
ID AAL47852 standard; DNA; 20775 BP.

AC AAL47852;

DT 19-SEP-2002 (first entry)

DE Murine basic helix-loop-helix BHLH-PAS protein related gene #2.

KW Basic helix-loop-helix protein; BHLH-PAS protein; neuroprotective;
KW neotrophic; Per-Arnt-Sim homology domain; developmental disease;
KW cell proliferative disease; nerve function; dementia; gene; ds;
KW Alzheimer's disease; memory.

OS Mus musculus.

PN WO200253729-A1.

PD 11-JUL-2002.

PF 17-DEC-2001; 2001WO-JP11064.

PR 27-DEC-2000; 2000JP-0398548.

PR 19-MAR-2001; 2001JP-0077740.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI One N;

DR WPI; 2002-528860/56.

PT Mammalian basic helix-loop-helix bHLH-PAS proteins and encoding genes
PT with transcriptional regulatory ability, useful in diagnosis and
PT developing drugs for e.g. genetic diseases related to nerve functions,
PT motion and long-term memory

PS Example 10; Page 157-173; 176pp; Japanese.

CC The present invention provides the protein and coding sequences of basic
CC helix-loop-helix Per-Arnt-Sim domain containing proteins (BHLH-PAS
CC proteins) from the human, mouse and rat. The sequences are useful in the
CC diagnosis and treatment of diseases associated with cell proliferation
CC and differentiation, development, functional expression, particularly
CC genetic diseases related to nerve function, motion and long-term memory
CC like dementia or Alzheimer's disease. The present sequence is a coding
CC sequence of the invention.

SO Sequence 20775 BP; 5384 A; 5246 C; 5100 G; 5045 T; 0 other;

Query Match 1.6%; Score 140.8; DB 24; Length 20775;
Best Local Similarity 83.3%; Pred. No. 1.8e-23;
Matches 160; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 4572 TTAAGAAAAAATGAAGCCAGCAGGTGGCAAGCCCTTTATCCAGACTTGGAG 4631
|||
DB 19538 TAGAAATGTAAATTTAGTCCGGGCGTGCTGCTGCCTTTATCCAGACTCGGAG 19597
OY 4632 GCAGAAACGACGAGATTTCGAGTTCAGGCGCAGCCTGCTATAGAGTAGTTCAGGA 4691

DB 19598 GCAGAGGCAAGGAGATTTCGAGTTCGAGGCCAGCCTGGTCTAACAAGTAGTTCAGGA 19657
OY 4692 CAGCCAGGGCTTACACAGAGAAACCTGTTTGA AAAACCAAGAAAAACAAACAA 4751
|||||
DB 19658 CAGCCAGGGCTTACACAGAGAAACCTGTCTCGAAGACAAACAAACAAACAA 19717
OY 4752 ACAAAACAAAAAC 4763
|||||
DB 19718 ACAAAACAAAAAC 19729

RESULT 13
ABN83212
ID ABN83212 standard; DNA; 20775 BP.

AC ABN83212;

DT 13-SEP-2002 (first entry)

DE Mouse transcriptional regulator gene #2.

KW Mouse; transcriptional regulator; neotrophic; neuroprotective; aging;
KW cognition failure; dementia; Alzheimer's disease; gene; ds.
KW Mus musculus.

FH Location/Qualifiers

FT CDS

FT /tag= a "Transcriptional regulator"

FT /note= "Contains 7 introns"

FT /tag= b

FT /number= 1

FT /tag= c

FT /number= 1

FT /tag= d

FT /number= 2

FT /tag= e

FT /number= 2

FT /tag= f

FT /number= 3

FT /tag= g

FT /number= 3

FT /tag= h

FT /number= 4

FT /tag= i

FT /number= 4

FT /tag= j

FT /number= 5

FT /tag= k

FT /number= 5

FT /tag= l

FT /number= 6

FT /tag= m

FT /number= 6

FT /tag= n

FT /number= 7

FT /tag= o

FT /number= 0

FT /tag= 0

FT /number= 0

FT /tag= 0

FT /number= 0

FT /tag= 0

FT /number= 0

FT /tag= 0

FT /number= 0

